А. В. КУЗНЕЦОВ,
В. А. САКОВИЧ, Н. И. ХОЛОД

ВЫСШАЯ МАТЕМАТИКА

МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

УЧЕБНИК

Под общей редакцией А. В. КУЗНЕЦОВА

Издание четвертое, стереотипное

ЛАНЬ® • САНКТ-ПЕТЕРБУРГ • МОСКВА • КРАСНОДАР • 2013

ББК 22.18я73
К 89

Кузнецов А. В., Сакович В. А., Холод Н. И.

К 89 Высшая математика. Математическое программирование / Под общ. ред. А. В. Кузнецова: Учебник. 4-е изд., стер. — СПб.: Издательство «Лань», 2013. — 352 с.: ил. — (Учебники для вузов. Специальная литература).

ISBN 978-5-8114-1056-9

Излагаются методы решения задач линейного программирования, элементы теории двойственности, рассматриваются программирование на сетях, дискретное и выпуклое программирование, основы теории матричных игр, динамического и параметрического программирования, даются сведения из стохастического программирования, излагаются методы решения задач транспортного типа. Основное внимание уделено приложениям математических методов в экономике, приведены примеры экономического содержания с анализом полученных результатов. С материалом книги согласован «Сборник задач и упражнений по высшей математике: математическое программирование» под ред. А. В. Кузнецова и Р. А. Рутковского.

Учебник предназначен для студентов вузов, обучающихся по экономическим направлениям подготовки.

ББК 22.18я73

Обложка
А. Ю. ЛАПШИН

ПРЕДИСЛОВИЕ

Предлагаемый учебник посвящен математическому программированию — области математики, разрабатывающей теорию и численные методы решения многомерных оптимизационных задач с ограничениями, т. е. задач на экстремум функции многих переменных с ограничениями на область изменения этих переменных. В отличие от классической теории экстремальных задач в математическом программировании основное внимание уделяется тем задачам, в которых активно участвуют ограничения на область изменения переменных. Создание методов математического программирования связано с насущными потребностями планирования и организации производства.

В учебнике излагаются теория и численные методы решения задач линейного программирования, элементы теории двойственности в линейном программировании, основы теории матричных игр, рассматривается программирование на сетях, отдельно излагаются методы решения задач транспортного типа, даются основы дискретного и выпуклого программирования, приводятся элементы динамического и параметрического программирования, даются сведения из стохастического программирования.

Во втором издании учебника расширена глава, посвященная теории двойственности, за счет включения материала о применении двойственных оценок в послеоптимизационном исследовании результатов решения линейных задач и общем анализе линейных моделей. В приложении рассматривается использование обратной матрицы при определении пределов чувствительности двойственных оценок.

При изучении математического программирования студенту потребуется знание общего курса высшей математики, теории вероятностей, математической статистики. Он должен свободно владеть математическим аппаратом, необходимым для решения теоретических и практических задач экономики и планирования. Кроме того, от него потребуются знания и навыки по программированию на ЭВМ, а также умение пользоваться персональными ЭВМ.

Поскольку выпускники вузов по экономическим специальностям в последующей практической деятельности будут встречаться с математическими методами оптимизации главным образом как пользователи, а не разработчики, в данном учебнике основное внимание уделяется приложениям математических методов в экономике, а не их подробному теоретическому обоснованию. По этой причине в учебнике приводится достаточное количество содержательных примеров, иллюстрирующих приемы математического моделирования экономических ситуаций с последующим экономическим анализом полученных результатов. Что же касается углубленного математического обоснования рассматриваемых в учебнике методов оптимизации, то это можно найти в специальной литературе, недостатка в которой в настоящее время не ощущается. При работе над книгой авторы использовали как отечественные, так и зарубежные источники, часть из которых указана в списке литературы.

Авторы выражают глубокую благодарность сотрудникам кафедры прикладной математики и экономической кибернетики Белорусского государственного экономического университета за объективное и заинтересованное обсуждение учебника, а также рецензенту — коллективу кафедры высшей математики № 2 Белорусской государственной политехнической академии, особенно ее заведующему, канд. физ.-мат. наук, доц. А. Д. Корзникову за ценные советы и замечания, способствовавшие улучшению книги.

Авторы

ВВЕДЕНИЕ

Предмет математического программирования. Многие задачи, с которыми приходится иметь дело в повседневной практике, являются многовариантными. Среди множества возможных вариантов в условиях рыночных отношений приходится отыскивать наилучшие в некотором смысле при ограничениях, налагаемых на природные, экономические и технологические возможности. До недавнего времени большинство таких задач решали исходя из здравого смысла и опыта лиц, принимающих решения, или просто "на глаз". При таком подходе не было и не могло быть никакой уверенности, что найденный вариант — наилучший. При современных масштабах производства даже незначительные ошибки оборачиваются громадными потерями. В связи с этим возникла необходимость применять для анализа и синтеза экономических ситуаций и систем математические методы и современную вычислительную технику. Такие методы объединяются под общим названием — математическое программирование.

Математическое программирование — область математики, разрабатывающая теорию и численные методы решения многомерных экстремальных задач с ограничениями, т. е. задач на экстремум функции многих переменных с ограничениями на область изменения этих переменных.

Функцию, экстремальное значение которой нужно найти в условиях экономических возможностей, называют *целевой, показателем эффективности* или *критерием оптимальности.* Экономические возможности формализуются в виде *системы ограничений.* Все это составляет математическую модель.

Математическая модель задачи — это отражение оригинала в виде функций, уравнений, неравенств, цифр и т. д. Модель задачи математического программирования включает:

1) совокупность неизвестных величин $\mathbf{x} = (x_1; \dots ; x_j; \dots ; x_n)$, действуя на которые, систему можно совершенствовать.

Их называют *планом задачи* (вектором управления, решением, управлением, стратегией, поведением и др.);

2) целевую функцию (функцию цели, показатель эффективности, критерий оптимальности, функционал задачи и др.). Целевая функция позволяет выбирать наилучший вариант из множества возможных. Наилучший вариант доставляет целевой функции экстремальное значение. Целевую функцию обозначим буквой Z $(Z = z(\mathbf{x}))$. Это может быть прибыль, объем выпуска или реализации, затраты производства, издержки обращения, уровень обслуживания или дефицитности, число комплектов, отходы и т. д.;

3) условия (или систему ограничений), налагаемые на неизвестные величины. Эти условия следуют из ограниченности ресурсов, которыми располагает общество в любой момент времени, из необходимости удовлетворения насущных потребностей, из условий производственных и технологических процессов. Ограниченными являются не только материальные, финансовые и трудовые ресурсы. Таковыми могут быть возможности технического, технологического и вообще научного потенциала. Нередко потребности превышают возможности их удовлетворения. Математически ограничения выражаются в виде уравнений и неравенств. Их совокупность образует *область допустимых решений (область экономических возможностей)*. Объединение всех условий (ограничений), налагаемых на неизвестные (искомые) величины x_j задачи, обозначим буквой Ω $(\mathbf{x} \in \Omega)$. При таких обозначениях модель задачи математического программирования примет вид $\max(\min)\ Z = z(\mathbf{x})$, $\mathbf{x} \in \Omega$, или: найти extremum $Z = z(\mathbf{x})$, $\mathbf{x} \in \Omega$.

В развернутом виде: найти план $\mathbf{x} = (x_1; \ldots; x_j; \ldots; x_n)$, доставляющий экстремальное значение целевой функции Z, т. е.

$$\max(\min)\ Z = z(x_1, \ldots, x_j, \ldots, x_n)$$

при ограничениях

$$\varphi_i(x_1, \ldots, x_j, \ldots, x_n) \{\leqslant, =, \geqslant\}\ b_i \quad (i = \overline{1, m}).$$

Из экономических или физических соображений на план задачи или некоторые его компоненты (координаты), как правило, налагаются условия неотрицательности

$$x_j \geqslant 0, \quad j \in \Omega_1 \subset \Omega,$$

иногда — целочисленности.

План $\mathbf{x}$, удовлетворяющий системе ограничений задачи, называется *допустимым* ($\mathbf{x} \in \Omega$). Допустимый план, доставляющий функции цели экстремальное значение, называется *оптимальным*. Оптимальный план будем обозначать $\mathbf{x}^*$, экстремальное значение функции цели — $z(\mathbf{x}^*) = Z^*$. Оптимальное решение, вообще говоря, не обязательно единственно, возможны случаи, когда оно не существует, имеется конечное или бесчисленное множество оптимальных решений.

Краткая классификация методов математического программирования. В зависимости от особенностей целевой функции $z(\mathbf{x})$ и функций, задающих ограничения $\varphi_i(\mathbf{x})$, задачи математического программирования делятся на ряд типов.

Если целевая функция $Z = z(\mathbf{x})$ и функции $\varphi_i(\mathbf{x})$ $(i = \overline{1, m})$, входящие в систему ограничений, линейны (первой степени) относительно входящих в задачу неизвестных x_j, то такой раздел математического программирования называется *линейным программированием* (ЛП). Методы и модели линейного программирования широко применяются при оптимизации процессов во всех отраслях народного хозяйства: при разработке производственной программы предприятия, распределении ее по исполнителям, при размещении заказов между исполнителями и по временны́м интервалам, при определении наилучшего ассортимента выпускаемой продукции, в задачах перспективного, текущего и оперативного планирования и управления; при планировании грузопотоков, определении плана товарооборота и его распределении; в задачах развития и размещения производительных сил, баз и складов систем обращения материальных ресурсов и т. д. Особенно широкое применение методы и модели линейного программирования получили при решении задач экономии ресурсов (выбор ресурсосберегающих технологий, составление смесей, раскрой материалов), производственно-транспортных и других задач.

Начало линейному программированию было положено в

1939 г. советским математиком-экономистом Л. В. Канторовичем в работе "Математические методы организации и планирования производства". Появление этой работы открыло новый этап в применении математики в экономике. Спустя десять лет американский математик Дж. Данциг разработал эффективный метод решения данного класса задач — симплекс-метод. Термин "линейное программирование" впервые появился в 1951 г. в работах Дж. Данцига и Т. Купманса.

Линейное программирование и межотраслевой баланс характеризуют линейные взаимосвязи элементов народного хозяйства.

Однако при более глубоком исследовании в ряде задач появляются и связи нелинейного характера, когда с изменением одного элемента другие изменяются непропорционально первому. Поэтому вслед за разработкой моделей линейного программирования начались интенсивные исследования нелинейных моделей.

Если в задаче математического программирования целевая функция $z(\mathbf{x})$ и (или) хотя бы одна из функций системы ограничений $\varphi_i(\mathbf{x})$ нелинейны, то такой раздел называется *нелинейным программированием* (НЛП). Методы и модели нелинейного программирования могут применяться при решении перечисленных выше задач, когда хотя бы одна из функций $z(\mathbf{x})$, $\varphi_i(\mathbf{x})$ нелинейна. Кроме того, методы НЛП получили широкое применение при расчете экономически выгодных партий запуска деталей в производство, при определении экономически выгодной партии поставки, поставочного комплекта, размеров запасов, распределении ограниченных ресурсов, размещении производительных сил, в тарном хозяйстве, при решении многих производственно-экономических задач и т. д.

Если на все или некоторые переменные x_j наложено условие дискретности, например целочисленности ($x_j = 0, 1, 2, \dots$), то такие задачи рассматриваются в разделе математического программирования, называемом дискретным, в частности *целочисленным* (ЦП), программированием. Методами ЦП решается широкий круг задач оптимизации с неделимостями, комбинаторного типа, с логическими условиями, с разрывной целевой функцией и т. д. В частности, задачи выбора (о назначениях), о контейнерных перевозках (о рюкзаке), о маршрутизации (коммивояжера), теории расписаний, комплектных поставок и комплектования, размещения производственно-складской структуры и т. п.

Если параметры целевой функции и (или) системы ограничений изменяются во времени или целевая функция имеет аддитивный

$$z(\mathbf{x}) = \sum_{j=1}^{n} z_j(\mathbf{x}_j),$$

либо мультипликативный

$$z(\mathbf{x}) = \prod_{j=1}^{n} z_j(\mathbf{x}_j)$$

вид или сам процесс выработки решения имеет многошаговый характер, то такие задачи решаются методами *динамического программирования* (ДП). Методами ДП могут решаться задачи перспективного и текущего планирования, управления производством, поставками и запасами в условиях изменяющегося спроса, распределения ограниченных ресурсов, в частности размещения капитальных вложений, замены оборудования, обновления и восстановления элементов сложных человеко-машинных организационных систем и т. д.

В перечисленных выше разделах математического программирования предполагается, что вся информация о протекании процессов заранее известна и достоверна. Такие методы оптимизации называются *детерминированными* или методами обоснования решений в условиях определенности.

Если параметры, входящие в функцию цели, или ограничения задачи являются случайными, недостоверными величинами или если приходится принимать решения в условиях риска, неполной или недостоверной информации, то говорят о проблеме стохастической оптимизации, а соответствующий раздел называется *стохастическим программированием* (СП). К нему в первую очередь следует отнести методы и модели выработки решений в условиях конфликтных ситуаций (математическая теория игр), при неполной информации (экспертные оценки), в условиях риска (статистические решения) и др. Позднее появились иные типы задач, учитывающих специфику целевой функции и системы ограничений, в связи с чем возникли параметрическое, дробно-линейное, блочное, сетевое (потоковое), многоиндексное, булевское, комбинаторное и другие типы программирования. В случае нелинейностей специфика задач породила квадратичное, биквадратичное, сепарабельное, выпуклое и другие типы программирования. Появились численные методы отыскания оптимальных решений: градиентные, штрафных и барьерных функций, возможных направлений, линейной аппроксимации, случайного поиска и др.

К математическому программированию относятся также методы решения экстремальных задач с бесконечным числом переменных — *бесконечномерное программирование.*

И, наконец, отметим, что задачи математического программирования с одной целевой функцией решаются методами скалярной оптимизации. Однако реальные ситуации настолько сложны, что нередко приходится одновременно учитывать несколько целевых функций, которые должны принимать экстремальные значения. Например, дать продукции больше, высокого качества и с минимальными затратами. Задачи, где находят решение по нескольким целевым функциям, относятся к векторной оптимизации — это так называемые *задачи многокритериального подхода.*

1. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

1.1. ПРИМЕРЫ ЭКОНОМИЧЕСКИХ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Понятие линейного программирования. Линейное программирование — раздел математического программирования, применяемый при разработке методов отыскания экстремума линейных функций нескольких переменных при линейных дополнительных ограничениях, налагаемых на переменные. По типу решаемых задач его методы разделяются на универсальные и специальные. С помощью универсальных методов могут решаться любые *задачи линейного программирования* (ЗЛП). Специальные методы учитывают особенности модели задачи, ее целевой функции и системы ограничений.

Особенностью задач линейного программирования является то, что экстремума целевая функция достигает на границе области допустимых решений. Классические же методы дифференциального исчисления связаны с нахождением экстремумов функции во внутренней точке области допустимых значений. Отсюда — необходимость разработки новых методов.

Задача о наилучшем использовании ресурсов. Пусть некоторая производственная единица (цех, завод, объединение и т. д.), исходя из конъюнктуры рынка, технических или технологических возможностей и имеющихся ресурсов, может выпускать n различных видов продукции (товаров), известных под номерами, обозначаемыми индексом j ($j = \overline{1, n}$). Ее будем обозначать Π_j. Предприятие при производстве этих видов продукции должно ограничиваться имеющимися видами ресурсов, технологий, других производственных факторов (сырья, полуфабрикатов, рабочей силы, оборудования, электроэнергии и т. д.). Все эти виды ограничивающих факторов называют *ингредиентами* R_i. Пусть их число равно m; припишем им индекс i ($i = \overline{1, m}$). Они ограничены, и их количества равны соответственно $b_1, \ldots, b_i, \ldots, b_m$

условных единиц. Таким образом, $\mathbf{b} = (b_1; \ldots; b_i; \ldots; b_m)$ — вектор ресурсов. Известна экономическая выгода (мера полезности) производства продукции каждого вида, исчисляемая, скажем, по отпускной цене товара, его прибыльности, издержкам производства, степени удовлетворения потребностей и т. д. Примем в качестве такой меры, например, цену реализации c_j $(j = \overline{1,n})$, т. е. $\mathbf{c} = (c_1; \ldots; c_j; \ldots; c_n)$ — вектор цен. Известны также технологические коэффициенты a_{ij}, которые указывают, сколько единиц i-го ресурса требуется для производства единицы продукции j-го вида. Матрицу коэффициентов a_{ij} называют *технологической* и обозначают буквой A. Имеем $A = [a_{ij}]$. Обозначим через $\mathbf{x} = (x_1; \ldots; x_j; \ldots; x_n)$ план производства, показывающий, какие виды товаров $\Pi_1, \ldots, \Pi_j, \ldots, \Pi_n$ нужно производить и в каких количествах, чтобы обеспечить предприятию максимум объема реализации при имеющихся ресурсах.

Так как c_j — цена реализации единицы j-й продукции, цена реализованных x_j единиц будет равна $c_j x_j$, а общий объем реализации

$$Z = c_1 x_1 + \ldots + c_n x_n.$$

Это выражение — целевая функция, которую нужно максимизировать.

Так как $a_{ij} x_j$ — расход i-го ресурса на производство x_j единиц j-й продукции, то, просуммировав расход i-го ресурса на выпуск всех n видов продукции, получим общий расход этого ресурса, который не должен превосходить b_i $(i = \overline{1,m})$ единиц:

$$a_{i1} x_1 + \ldots + a_{ij} x_j + \ldots + a_{in} x_n \leqslant b_i.$$

Чтобы искомый план $\mathbf{x} = (x_1; \ldots; x_j; \ldots; x_n)$ был реален, наряду с ограничениями на ресурсы нужно наложить условие неотрицательности на объемы x_j выпуска продукции:

$$x_j \geqslant 0 \quad (j = \overline{1,n}).$$

Таким образом, модель задачи о наилучшем использовании ресурсов примет вид: найти

$$\max Z = \sum_{j=1}^{n} c_j x_j \tag{1.1}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \quad (i = \overline{1, m}), \tag{1.2}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}). \tag{1.3}$$

Так как переменные x_j входят в функцию $z(\mathbf{x})$ и систему ограничений только в первой степени, а показатели a_{ij}, b_i, c_j являются постоянными в планируемый период, то (1.1) – (1.3) — задача линейного программирования.

Задача о выборе оптимальных технологий. В задаче о наилучшем использовании ресурсов определяется оптимальный план выпуска продукции. Пусть при производстве какого-то общественно необходимого продукта используется n технологий. При этом требуется m видов ресурсов, заданных объемами b_i $(i = \overline{1, m})$. Эффективности технологий, т. е. количество конечной продукции (в ден. ед.), производимой в единицу времени по j-й $(j = \overline{1, n})$ технологии, обозначим c_j. Пусть, далее, a_{ij} — расход i-го ресурса в единицу времени по j-й технологии. В качестве неизвестной величины x_j примем интенсивность использования j-й технологии, т. е. время, в течение которого продукция производится по j-й технологии. Пренебрегая временем переналадок, необходимым для перехода от одной технологии к другой, получаем следующую математическую модель задачи: найти план интенсивностей использования технологий $\mathbf{x} = (x_1; \ldots ; x_n)$, обеспечивающий максимум выпуска продукции в стоимостном выражении:

$$\max Z = \sum_{j=1}^{n} c_j x_j \tag{1.4}$$

при ограничениях на лимитируемые ресурсы

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \quad (i = \overline{1, m}) \tag{1.5}$$

и условии неотрицательности

$$x_j \geqslant 0 \quad (j = \overline{1, n}). \tag{1.6}$$

Задача о смесях. В различных отраслях народного хозяйства возникает проблема составления таких рабочих смесей на основе исходных материалов, которые обеспечивали бы получение конечного продукта, обладающего определенными свойствами. К этой группе задач относятся задачи о выборе диеты, составлении кормового рациона в животноводстве, шихт в металлургии, горючих и смазочных смесей в нефтеперерабатывающей промышленности, смесей для получения бетона в строительстве и т. д. Высокий уровень затрат на исходные сырьевые материалы и необходимость повышения эффективности производства выдвигают на первый план следующую задачу: получить продукцию с заданными свойствами при наименьших затратах на исходные сырьевые материалы.

Модель задачи о наилучшем составе смеси рассмотрим на примере задачи о диете. Имеются пищевые продукты, известные под номерами $1, 2, \ldots, n$. Они содержат различные питательные вещества, обозначаемые номерами $1, 2, \ldots, m$ (углеводы, белки, жиры, витамины, микроэлементы и др.). Единица j-го продукта содержит a_{ij} единиц i-го питательного вещества. Для нормальной жизнедеятельности в заданный промежуток времени нужно потреблять не менее b_i единиц i-го питательного вещества. Обозначим через c_j стоимость единицы продукта j-го вида. Требуется выбрать рацион минимальной стоимости, содержащий необходимые количества питательных веществ. План задачи — это количества x_j продуктов каждого вида, обеспечивающие необходимое количество питательных веществ при минимальных затратах на исходные продукты.

Математическая модель задачи: найти

$$\min Z = \sum_{j=1}^{n} c_j x_j \tag{1.7}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad (i = \overline{1, m}), \tag{1.8}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}). \tag{1.9}$$

Задача о раскрое материалов. Суть задачи об оптимальном раскрое состоит в разработке таких технологически допустимых планов раскроя, при которых получается необходимый комплект заготовок, а отходы (по длине, площади, объему, массе или стоимости) сводятся к минимуму. Рассмотрим простейшую модель раскроя по одному измерению. Более сложные постановки ведут к задачам целочисленного программирования.

Модель задачи раскроя по одному измерению длинномерных материалов (прутков, труб, профильного проката и др.) может быть сформулирована так. Пусть имеется N штук исходного материала, длина каждой штуки равна L. Нужны заготовки m видов, длины которых равны l_i $(i = \overline{1,m})$. Известна потребность в заготовках каждого вида, она равна b_i. Изучение вопроса раскроя (построение технологической карты раскроя) показывает, что можно выделить n приемлемых вариантов раскроя исходного материала длиной L на заготовки длиной l_i. Обозначим через a_{ij} количество заготовок i-го вида, получаемое при раскрое единицы исходного материала по j-му $(j = \overline{1,n})$ варианту, c_j — отходы при раскрое единицы исходного материала по j-му варианту. План задачи $\mathbf{x} = (x_1; \ldots; x_j; \ldots; x_n)$, где x_j — количество единиц исходного материала, планируемое к раскрою по j-му варианту.

Функция цели — минимум отходов, получаемых при раскрое:

$$\min Z = \sum_{j=1}^{n} c_j x_j \tag{1.10}$$

при ограничениях: на число единиц исходного материала

$$\sum_{j=1}^{n} x_j \leqslant N, \tag{1.11}$$

на удовлетворение ассортиментного спроса потребителей

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad (i = \overline{1,m}) \tag{1.12}$$

и условии неотрицательности

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{1.13}$$

Транспортная задача. Рассмотрим простейший вариант модели транспортной задачи, когда речь идет о рациональной перевозке некоторого однородного продукта от производителей к потребителям; при этом имеется баланс между суммарным спросом потребителей и возможностями поставщиков по их удовлетворению. Причем потребителям безразлично, из каких пунктов производства будет поступать продукция, лишь бы их заявки были полностью удовлетворены. Так как от схемы прикрепления потребителей к поставщикам существенно зависит объем транспортной работы, возникает задача о наиболее рациональном прикреплении, правильном направлении перевозок грузов, при котором потребности полностью удовлетворяются, вся продукция от поставщиков вывозится, а затраты на транспортировку минимальны.

Задача формулируется так. Имеется m пунктов производства, в каждом из которых сосредоточено a_i $(i = \overline{1, m})$ единиц однородного продукта. Этот продукт нужно доставить n потребителям, где потребность составляет b_j $(j = \overline{1, n})$ единиц. Причем $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$.

Известны величины c_{ij} — затраты на перевозку единицы продукта из i-го пункта производства в j-й пункт потребления. Обозначим через x_{ij} количество продукта, перевозимое из i-го пункта производства в j-й пункт потребления. Матрица $C = [c_{ij}]$ называется *матрицей тарифов*, $X = [x_{ij}]$ — *матрицей перевозок*. С целью удобства построения математической модели матрицы тарифов и перевозок совмещают в одну, именуемую *макетом транспортной задачи* (табл. 1.1).

Математическая модель транспортной задачи: целевая функция, описывающая транспортные затраты,

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{1.14}$$

минимизируется при ограничениях: на возможности поставщиков — весь продукт из пунктов производства должен быть вывезен:

$$\sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1, m}), \tag{1.15}$$

на спрос потребителей, который должен быть удовлетворен:

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1,n}) \tag{1.16}$$

при условии неотрицательности переменных, исключающем обратные перевозки:

$$x_{ij} = 0 \quad (i = \overline{1,m};\ j = \overline{1,n}). \tag{1.17}$$

Таблица 1.1

a_i \ b_j	b_1	b_2	...	b_j	...	b_n
a_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1j} x_{1j}	...	c_{1n} x_{1n}
a_2	c_{21} x_{21}	c_{22} x_{22}	...	c_{2j} x_{2j}	...	c_{2n} x_{2n}
...	...	...	...	...	...	...
a_i	c_{i1} x_{i1}	c_{i2} x_{i2}	...	c_{ij} x_{ij}	...	c_{in} x_{in}
...	...	...	...	...	...	...
a_m	c_{m1} x_{m1}	c_{m2} x_{m2}	...	c_{mj} x_{mj}	...	c_{mn} x_{mn}

Задача о размещении заказа. Речь идет о задаче распределения заказа (загрузки взаимозаменяемых групп оборудования) между предприятиями (цехами, станками, исполнителями) с различными производственными и технологическими характеристиками, но взаимозаменяемыми в смысле выполнения заказа. Требуется составить план размещения заказа (загрузки оборудования), при котором с имеющимися производственными возможностями заказ был бы выполнен, а показатель эффективности достигал экстремального значения.

Сформулируем задачу конкретнее. Имеется m однородных групп оборудования, на котором нужно выполнить заказ на выпуск n видов продукции в объемах x_j^* $(j = \overline{1,n})$ единиц.

Заказ определяется набором x_j^* $(j = \overline{1, n})$, который устанавливается решением задачи о наилучшем использовании ресурсов, изучением структуры потребления или просто "спущен сверху". Мощность оборудования каждого вида ограничена, например, фондом рабочего времени T_i $(i = \overline{1, m})$. Производительность оборудования каждого вида задана коэффициентом a_{ij}, который показывает, сколько единиц продукции j-го вида можно произвести на i-м оборудовании в единицу времени. Кроме того, известны коэффициенты c_{ij}, отражающие все затраты, вызванные изготовлением на i-м оборудовании в единицу времени продукции j-го вида. Требуется найти план $X = [x_{ij}]$ размещения заказа (загрузки оборудования), т. е. установить, сколько времени i-я группа оборудования будет занята изготовлением j-й продукции.

Целевая функция (суммарные затраты на выполнение заказа)

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{1.18}$$

минимизируется при нижеследующих ограничениях. По мощности оборудования

$$\sum_{j=1}^{n} x_{ij} \leqslant T_i \quad (i = \overline{1, m}). \tag{1.19}$$

Если по некоторым видам продукции допускается перевыполнение плана, то

$$\sum_{i=1}^{m} a_{ij} x_{ij} \geqslant x_j^* \quad (j = \overline{1, n_1}). \tag{1.20}$$

Для продукции, выпуск которой должен соответствовать плану,

$$\sum_{i=1}^{m} a_{ij} x_{ij} = x_j^* \quad (j = \overline{n_1 + 1, n_2}). \tag{1.21}$$

Для продукции, заказ на которую принимается для более полной загрузки оборудования,

$$\sum_{i=1}^{m} a_{ij} x_{ij} \leqslant x_j^* \quad (j = \overline{n_2 + 1, n}). \tag{1.22}$$

Условие неотрицательности следует из практического смысла переменных:

$$x_{ij} \geqslant 0 \quad (i = \overline{1,m};\ j = \overline{1,n}\,). \tag{1.23}$$

Задачи (1.1) – (1.3), (1.4) – (1.6), (1.7) – (1.9), (1.10) – (1.13), (1.14) – (1.17), (1.18) – (1.23) относятся к ЗЛП.

1.2. ФОРМЫ ЗАПИСИ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ, ИХ ЭКВИВАЛЕНТНОСТЬ И СПОСОБЫ ПРЕОБРАЗОВАНИЯ

Основные виды записи ЗЛП. *Общей задачей линейного программирования* (ОЗЛП) называют задачу

$$\max(\min) Z = \sum_{j=1}^{n} c_j x_j \tag{1.24}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1,m_1}\,), \tag{1.25}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = \overline{m_1+1,m_2}\,), \tag{1.26}$$

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad (i = \overline{m_2+1,m}\,), \tag{1.27}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n_1}\,), \tag{1.28}$$

$$x_j \text{— произвольные } (j = \overline{n_1+1,n}\,), \tag{1.29}$$

где c_j, a_{ij}, b_i — заданные действительные числа; (1.24) — целевая функция; (1.25) - (1.29) — ограничения; $\mathbf{x} = (x_1; \ldots ; x_n)$ — план задачи.

Симметричной формой записи ЗЛП называют задачу

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{1.30}$$

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \quad (i = \overline{1,m}), \tag{1.31}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}) \tag{1.32}$$

или задачу

$$\min Z = \sum_{j=1}^{n} c_j x_j; \tag{1.33}$$

$$\sum_{j=1}^{n} a_{ij}x_j \geqslant b_i \quad (i = \overline{1,m}), \tag{1.34}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{1.35}$$

В экономической практике задача (1.30) — (1.32) (или (1.33) — (1.35)) встречается наиболее часто.

Канонической формой записи ЗЛП называют задачу

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{1.36}$$

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = \overline{1,m}), \tag{1.37}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{1.38}$$

Рассмотрим еще два употребительных вида записи — *матричную* и *векторную*. В модель (1.36) — (1.38) введем обозначения:

$$C = [\,c_1 \;\; c_2 \;\; \ldots \;\; c_n\,], \qquad A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix},$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix}, \qquad A_0 = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{bmatrix},$$

где C — матрица-строка; A — матрица системы уравнений; X — матрица-столбец переменных; A_0 — матрица-столбец свободных членов.

Каноническая форма задачи примет вид:

$$\max Z = [\,c_1 \;\; c_2 \;\; \dots \;\; c_n\,][\,x_1 \;\; x_2 \;\; \dots \;\; x_n\,]^T;$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}, \quad X \geqslant O,$$

или

$$\max Z = CX, \quad AX = A_0, \quad X \geqslant O.$$

Полезной является также векторная форма ЗЛП. Для столбцов матрицы A введем обозначения:

$$A_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{bmatrix}, \; A_2 = \begin{bmatrix} a_{21} \\ a_{22} \\ \dots \\ a_{m2} \end{bmatrix}, \dots, A_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{bmatrix}, \dots, A_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{bmatrix}.$$

Тогда задача (1.36) — (1.38) в векторной форме записи примет вид:

$$\max Z = \mathbf{cx};$$

$$A_1x_1 + \ldots + A_jx_j + \ldots + A_nx_n = A_0, \quad \mathbf{x} \geqslant \mathbf{0},$$

где $\mathbf{cx}$ — скалярное произведение векторов $\mathbf{c} = (c_1; \ldots; c_n)$ и $\mathbf{x} = (x_1; \ldots; x_n)$.

Способы преобразования. При необходимости задачу минимизации можно заменить задачей максимизации, и наоборот. Для функции одной переменной это утверждение очевидно. В самом деле, если x^* — точка минимума функции $y = f(x)$, то для функции $y = -f(x)$ она является точкой максимума, так как графики функций $f(x)$ и $-f(x)$ симметричны относительно оси абсцисс (рис. 1.1).

Итак,

$$\min f(x^*) = -\max\,(-f(x^*)).$$

То же самое имеет место и в случае функции n переменных:

$$\min f(x_1^*, \ldots, x_n^*) = -\max\,(-f(x_1^*, \ldots, x_n^*)).$$

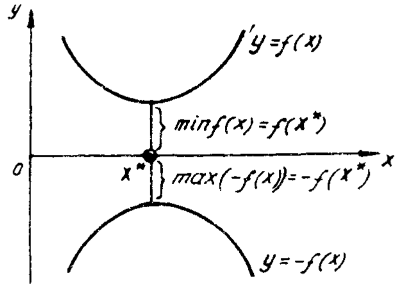


Рис. 1.1

Переход к канонической форме. Как следует из примеров задач линейного программирования (1.1) — (1.3), (1.4) — (1.6), (1.7) — (1.9), (1.10) — (1.13), (1.14) — (1.17), (1.18) — (1.23), в них большинство ограничений задается неравенствами. Наиболее же широко используемые методы решения ЗЛП применяются лишь к задачам, записанным в канонической форме. Поэтому приходится переходить от любой формы ЗЛП к ее каноническому виду, причем нужно быть уверенным, что эти формы эквивалентны.

Пусть исходная ЗЛП имеет вид:

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{1.39}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1, m_1}\,), \tag{1.40}$$

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad (i = \overline{m_1 + 1, m}\,), \tag{1.41}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}\,). \tag{1.42}$$

Преобразуем ее к каноническому виду. Введем m дополнительных (балансовых) неотрицательных переменных $x_{n+i} \geqslant 0$ $(i = \overline{1, m}\,)$. Для того чтобы неравенства типа $\leqslant$ (1.40) преобразовать в равенства, к их левым частям прибавим дополнительные переменные $x_{n+i} \geqslant 0$ $(i = \overline{1, m_1}\,)$, после чего система неравенств (1.40) примет вид

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \quad (i = \overline{1, m_1}). \tag{1.43}$$

Для того чтобы неравенства типа $\geqslant$ (1.41) преобразовать в равенства, из их левых частей вычтем дополнительные переменные x_{n+i} $(i = \overline{m_1 + 1, m})$. Получим

$$\sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i \quad (i = \overline{m_1 + 1, m}). \tag{1.44}$$

Систему уравнений (1.43) ((1.44)) с условием неотрицательности дополнительных переменных называют *эквивалентной* системе неравенств (1.40) ((1.41)).

Дополнительные переменные x_{n+i} $(i = \overline{1, m})$ в целевую функцию вводятся с коэффициентами, равными нулю. Получим задачу:

$$\max Z = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} 0 \cdot x_{n+i}; \tag{1.45}$$

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \quad (i = \overline{1, m_1}), \tag{1.46}$$

$$\sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i \quad (i = \overline{m_1 + 1, m}), \tag{1.47}$$

$$x_j \geqslant 0 \ (j = \overline{1, n}), \quad x_{n+i} \geqslant 0 \ (i = \overline{1, m}). \tag{1.48}$$

Задача (1.45) — (1.48) имеет каноническую форму. Задачи (1.39) — (1.42) и (1.45) — (1.48) тесно связаны между собой.

Теорема 1.1. *Каждому допустимому решению* $(x_1^0; \dots ; x_n^0)$ *задачи (1.39) — (1.42) соответствует вполне определенное допустимое решение* $(x_1^0; \dots ; x_n^0; x_{n+1}^0; \dots ; x_{n+m}^0)$ *задачи (1.45) — (1.48), где* $x_{n+i} \geqslant 0$ $(i = \overline{1, m})$, *и наоборот, каждому допустимому решению* $(x_1^0; \dots ; x_n^0; x_{n+1}^0; \dots ; x_{n+m}^0)$ *задачи (1.45) — (1.48) соответствует допустимое решение* $(x_1^0; \dots ; x_n^0)$ *задачи (1.39) — (1.42).*

Доказательство. Пусть $(x_1^0; \ldots ; x_n^0)$ — допустимое решение задачи (1.39) — (1.42). Для условий (1.40) обозначим

$$x_{n+i}^0 = b_i - \sum_{j=1}^{n} a_{ij} x_j^0 \geqslant 0, \tag{1.49}$$

для условий (1.41) —

$$x_{n+i}^0 = \sum_{j=1}^{n} a_{ij} x_j^0 - b_i \geqslant 0. \tag{1.50}$$

Из условий (1.49) и (1.50) следуют условия (1.46) — (1.48). Отсюда $\mathbf{x}^0 = (x_1^0; \ldots ; x_n^0; x_{n+1}^0; \ldots ; x_{n+m}^0)$ есть определенное допустимое решение задачи (1.45) — (1.48). △*

Аналогично доказывается обратное утверждение. Так как дополнительные переменные x_{n+i}^0 входят в целевую функцию (1.45) с коэффициентами, равными нулю, то значения целевых функций (1.39) и (1.45) при соответствующих допустимых решениях $(x_1^0; \ldots ; x_n^0)$ и $(x_1^0; \ldots ; x_n^0; x_{n+1}^0; \ldots ; x_{n+m}^0)$ одинаковы. Отсюда следует, что целевые функции (1.39) и (1.45) на множестве соответствующих допустимых решений достигают экстремального значения одновременно. Оптимальному решению $(x_1^*; \ldots ; x_n^*)$ задачи (1.39) — (1.42) соответствует оптимальное решение $(x_1^*; \ldots ; x_n^*; x_{n+1}^*; \ldots ; x_{n+m}^*)$ задачи (1.45) — (1.48), т. е. исходная задача и ее каноническая форма эквивалентны.

Пример 1.1. Магазин оптовой торговли реализует три вида продукции $П_1$, $П_2$ и $П_3$. Для этого используются два ограниченных ресурса — полезная площадь помещений, которая с учетом коэффициента оборачиваемости составляет 450 м^2, и рабочее время работников магазина — 600 чел.-ч. Товарооборот должен быть не меньше 240 тыс. ден. ед. Необходимо разработать план товарооборота, доставляющий максимум прибыли. Затраты ресурсов на реализацию и получаемая при этом прибыль представлены в табл. 1.2. Составить математическую модель задачи и привести ее к каноническому виду.

Решение. Обозначим через x_1, x_2, x_3 объемы продукции (в тыс. ден. ед.), подлежащей реализации. Модель задачи примет вид:

$$\max Z = 50x_1 + 65x_2 + 70x_3;$$

*Знак △ означает конец доказательства.

$$\left.\begin{aligned} 1,5x_1+2x_2+\ \ \ 3x_3 &\leqslant 450,\\ 3x_1+2x_2+1,5x_3 &\leqslant 600,\\ x_1+\ \ x_2+\ \ \ \ \ \ x_3 &\geqslant 240, \end{aligned}\right\}$$

$$x_j \geqslant 0 \quad (j=\overline{1,3}).$$

Таблица 1.2

Ресурсы	Затраты ресурсов на реализацию, тыс. ден. ед.			Объем ресурса
	$П_1$	$П_2$	$П_3$	
Полезная площадь, м^2	1,5	2	3	450
Рабочее время, чел.-ч	3	2	1,5	600
Прибыль, тыс. ден. ед.	50	65	70	

Перейдем к задаче в каноническом виде. Введем дополнительные переменные x_4, x_5, x_6, первые две из которых прибавим к левым частям первых двух неравенств, а третью вычтем из левой части третьего неравенства. В целевую функцию все дополнительные переменные введем с коэффициентом, равным нулю. Получим каноническую форму задачи:

$$\max Z = 50x_1 + 65x_2 + 70x_3 + 0\cdot x_4 + 0\cdot x_5 + 0\cdot x_6;$$

$$\left.\begin{aligned} 1,5x_1+2x_2+\ \ \ 3x_3+x_4\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ &=450,\\ 3x_1+2x_2+1,5x_3\ \ \ \ \ \ \ +x_5\ \ \ \ \ \ \ \ &=600,\\ x_1+\ \ x_2+\ \ \ \ \ \ x_3\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ -x_6 &=240, \end{aligned}\right\}$$

$$x_j \geqslant 0 \quad (j=\overline{1,6}).$$

Отметим экономический смысл дополнительных переменных. Они в каждой задаче прямо связаны с ее экономическим содержанием.

Например, для задачи (1.1) — (1.3) о наилучшем использовании ресурсов

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij}x_j \quad (i=\overline{1,m}),$$

т. е. дополнительная переменная показывает величину неиспользованного ресурса. Для задачи о смесях (1.7) — (1.9)

$$x_{n+i} = \sum_{j=1}^{n} a_{ij}x_j - b_i \quad (i=\overline{1,m}),$$

т. е. дополнительная переменная показывает потребление соответствующего питательного вещества в оптимальном плане сверх нормы.

В ряде производственно-экономических ситуаций не на все переменные налагаются условия неотрицательности. В подобных ситуациях, даже если ограничения представлены в виде равенств, задача не будет канонической. Для представления такой задачи в каноническом виде каждую из переменных x_k, на которые не наложено условие неотрицательности, заменим разностью двух неотрицательных переменных x'_k и x''_k, т. е. $x_k = x'_k - x''_k$, где $x'_k \geqslant 0$, $x''_k \geqslant 0$.

Переход к симметричной форме записи можно осуществить двумя способами.

Первый способ. Пусть в ОЗЛП имеются ограничения равенства

$$\sum_{j=1}^{n} a_{ij}x_j = b_i.$$

Каждое такое ограничение-равенство эквивалентно системе неравенств

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i, \quad \sum_{j=1}^{n} a_{ij}x_j \geqslant b_i.$$

Неравенство вида $\geqslant$ умножением обеих частей на -1 превращается в неравенство вида $\leqslant$, и наоборот.

Второй способ. Рассмотрим ЗЛП в каноническом виде (1.36) — (1.38). Приведем ее к симметричной форме. Пусть ранг системы (1.37) равен r и $r < n$. Тогда система будет иметь бесконечное множество решений. Не ограничивая общности, можно считать, что в матрице системы линейно независимы первые r столбцов. Например, методом исключений Гаусса систему преобразуем к виду

$$x_i + \sum_{j=r+1}^{n} \alpha_{ij}x_j = \beta_i \quad (i = \overline{1,r}). \tag{1.51}$$

Переменные $x_1, \ldots, x_r$ называют *базисными*, а $x_{r+1}, \ldots, x_n$ — *свободными*. Выразим целевую функцию через свободные переменные. Для этого подставим значения x_i из равенств

(1.51) в равенство (1.36) и получим

$$Z = \sum_{j=1}^{n} c_j x_j = c_1 x_1 + c_2 x_2 + \ldots + c_r x_r + c_{r+1} x_{r+1} +$$

$$+ c_{r+2} x_{r+2} + \ldots + c_n x_n = c_1 \Big(\beta_1 - \sum_{j=r+1}^{n} \alpha_{1j} x_j\Big) +$$

$$+ c_2 \Big(\beta_2 - \sum_{j=r+1}^{n} \alpha_{2j} x_j\Big) + \ldots + c_r \Big(\beta_r - \sum_{j=r+1}^{n} \alpha_{rj} x_j\Big) +$$

$$+ c_{r+1} x_{r+1} + c_{r+2} x_{r+2} + \ldots + c_n x_n = (c_1 \beta_1 + c_2 \beta_2 + \ldots +$$

$$+ c_r \beta_r) - \big(((c_1 \alpha_{1,r+1} + c_2 \alpha_{2,r+1} + \ldots + c_r \alpha_{r,r+1}) -$$

$$- c_{r+1}) x_{r+1} + ((c_1 \alpha_{1,r+2} + c_2 \alpha_{2,r+2} + \ldots + c_r \alpha_{r,r+2}) -$$

$$- c_{r+2}) x_{r+2} + \ldots + ((c_1 \alpha_{1n} + c_2 \alpha_{2n} + \ldots + c_r \alpha_{rn}) - c_n) x_n\big).$$

Введем обозначения: $\mathbf{c}_\text{Б}$ — вектор коэффициентов целевой функции, стоящих при базисных переменных, $\mathbf{A}_0$ — вектор свободных членов в равенствах (1.51), $\mathbf{A}_{r+k}$ — вектор коэффициентов при свободных переменных, т. е.

$$\Delta_0 = c_1 \beta_1 + c_2 \beta_2 + \ldots + c_r \beta_r = \mathbf{c}_\text{Б} \mathbf{A}_0,$$

$$\Delta_{r+1} = (c_1 \alpha_{1,r+1} + c_2 \alpha_{2,r+1} + \ldots + c_r \alpha_{r,r+1}) - c_{r+1} =$$

$$= \mathbf{c}_\text{Б} \mathbf{A}_{r+1} - c_{r+1},$$

$$\Delta_{r+2} = (c_1 \alpha_{1,r+2} + c_2 \alpha_{2,r+2} + \ldots + c_r \alpha_{r,r+2}) - c_{r+2} =$$

$$= \mathbf{c}_\text{Б} \mathbf{A}_{r+2} - c_{r+2},$$

. . . .

$$\Delta_n = (c_1 \alpha_{1n} + c_2 \alpha_{2n} + \ldots + c_r \alpha_{rn}) - c_n = \mathbf{c}_\text{Б} \mathbf{A}_n - c_n.$$

Используя эти обозначения, получаем

$$Z = \Delta_0 - \sum_{j=r+1}^{n} \Delta_j x_j.$$

Так как $x_j \geqslant 0$, из равенств (1.51) имеем

$$\sum_{j=r+1}^{n} \alpha_{ij} x_j \leqslant \beta_i \quad (i = \overline{1, r}).$$

Модель ЗЛП принимает вид:

$$\max Z = \Delta_0 - \sum_{j=r+1}^{n} \Delta_j x_j;$$

$$\sum_{j=r+1}^{n} \alpha_{ij} x_j \leqslant \beta_i \quad (i = \overline{1, r}),$$

$$x_j \geqslant 0 \quad (j = \overline{r+1, n}).$$

Отметим попутно, что в любом случае при подстановке базисных переменных в целевую функцию справедлива формула

$$Z = \Delta_0 - \sum_{j=1}^{n} \Delta_j x_j, \tag{1.52}$$

где

$$\Delta_0 = \mathbf{c}_\text{Б} \mathbf{A}_0, \quad \Delta_j = \mathbf{c}_\text{Б} \mathbf{A}_j - c_j \quad (j = \overline{1, n}). \tag{1.53}$$

Формулы (1.52), (1.53) используют для контроля вычислений при решении ЗЛП симплексным методом.

1.3. ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ И ГРАФИЧЕСКОЕ РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Геометрическая интерпретация экономических задач дает возможность наглядно представить их структуру, выявить особенности и открывает пути исследования более сложных свойств. ЗЛП с двумя переменными всегда можно решить графически. Однако уже в трехмерном пространстве такое решение усложняется, а в пространствах, размерность которых больше трех, графическое решение, вообще говоря, невозможно.

Случай двух переменных не имеет особого практического значения, однако его рассмотрение проясняет свойства ОЗЛП, приводит к идее ее решения, делает геометрически наглядными способы решения и пути их практической реализации.

Пусть дана задача

$$\max Z = c_1x_1 + c_2x_2; \tag{1.54}$$

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 \leqslant b_1, \\ a_{21}x_1 + a_{22}x_2 \leqslant b_2, \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 \leqslant b_m, \end{array}\right\} \tag{1.55}$$

$$x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{1.56}$$

Дадим геометрическую интерпретацию элементов этой задачи. Каждое из ограничений (1.55), (1.56) задает на плоскости x_1Ox_2 некоторую полуплоскость. Полуплоскость — выпуклое множество. Но пересечение любого числа выпуклых множеств является выпуклым множеством. Отсюда следует, что область допустимых решений задачи (1.54) — (1.56) есть выпуклое множество.

Напомним, что *выпуклым* называют множество, которое вместе с любыми своими точками $\mathbf{x}_1$ и $\mathbf{x}_2$ содержит и все точки $\mathbf{x}$ отрезка $[\mathbf{x}_1; \mathbf{x}_2]$, т. е. точки $\mathbf{x} = \lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$, где $0 \leqslant \lambda \leqslant 1$, являющиеся выпуклыми линейными комбинациями точек $\mathbf{x}_1$ и $\mathbf{x}_2$. Иногда это свойство записывают иначе: $\mathbf{x} = \lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$, где $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$; $\lambda_1 + \lambda_2 = 1$.

На рис. 1.2 представлены возможные ситуации, когда область допустимых решений ЗЛП — выпуклый многоугольник (*а*), неограниченная выпуклая многоугольная область (*б*), единственная точка (*в*), прямая линия (*г*), луч (*д*), отрезок (*е*), пустое множество (*ж*).

Перейдем к геометрической интерпретации целевой функции. Пусть область допустимых решений ЗЛП — непустое множество, например многоугольник $A_1A_2A_3A_4A_5A_6$ (рис. 1.3). Выберем произвольное значение целевой функции $Z = Z_0$. Получим $c_1x_1 + c_2x_2 = Z_0$. Это уравнение прямой линии. В точках прямой NM целевая функция сохраняет одно и то же постоянное значение Z_0. Считая в равенстве (1.54)

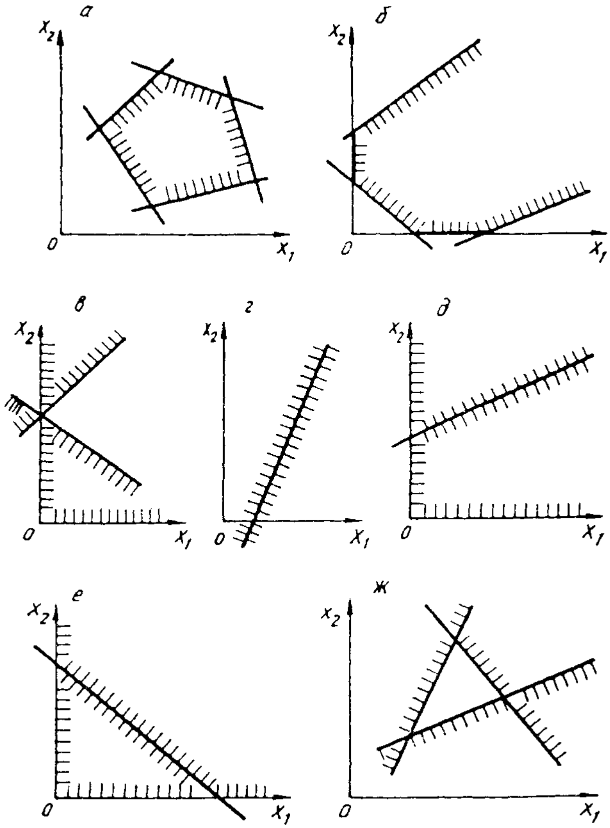


Рис. 1.2

Z параметром, получим уравнение семейства параллельных прямых, называемых *линиями уровня целевой функции* (*линиями постоянного значения*).

Возникает вопрос: как установить направление возрастания (убывания) целевой функции? Найдем частные производные целевой функции по x_1 и x_2:

$$\frac{\partial Z}{\partial x_1} = c_1, \tag{1.57}$$

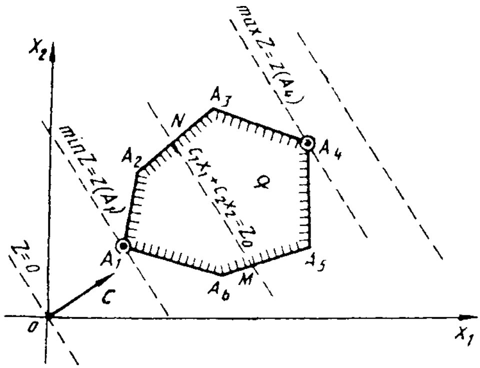


Рис. 1.3

$$\frac{\partial Z}{\partial x_2} = c_2. \tag{1.58}$$

Частная производная (1.57) ((1.58)) функции показывает скорость ее возрастания вдоль данной оси. Следовательно, c_1 и c_2 — скорости возрастания Z соответственно вдоль осей Ox_1 и Ox_2. Вектор $\mathbf{c} = (c_1; c_2)$ называется *градиентом функции.* Он показывает направление наискорейшего возрастания целевой функции:

$$\mathbf{c} = (\partial Z/\partial x_1, \partial Z/\partial x_2).$$

Вектор $-\mathbf{c}$ указывает направление наискорейшего убывания целевой функции. Его называют *антиградиентом.*

Вектор $\mathbf{c} = (c_1; c_2)$ перпендикулярен к прямым $Z = \text{const}$ семейства $c_1x_1 + c_2x_2 = Z$.

Из геометрической интерпретации элементов ЗЛП следует порядок ее графического решения.

1. С учетом системы ограничений строим область допустимых решений Ω.

2. Строим вектор $\mathbf{c} = (c_1; c_2)$ наискорейшего возрастания целевой функции — вектор градиентного направления.

3. Проводим произвольную линию уровня $Z = Z_0$ (проще всего провести линию $Z = 0$, перпендикулярную к вектору $\mathbf{c}$).

4. При решении задачи на максимум перемещаем линию уровня $Z = Z_0$ в направлении вектора $\mathbf{c}$ так, чтобы она касалась области допустимых решений в ее крайнем положении

(крайней точке) (на рис. 1.3 — до точки A_4). В случае решения задачи на минимум линию уровня $Z = Z_0$ перемещаем в антиградиентном направлении (на рис. 1.3 — до точки A_1).

5. Определяем оптимальный план $\mathbf{x}^* = (x_1^*; x_2^*)$ и экстремальное значение целевой функции $Z^* = z(\mathbf{x}^*)$.

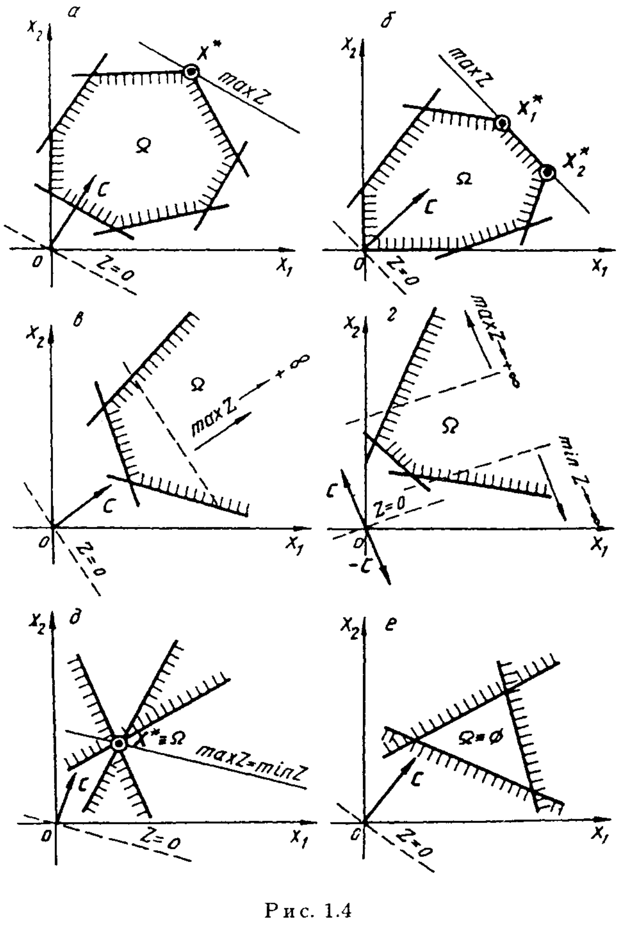


Рис. 1.4

Как видно из рис. 1.4, возможны следующие случаи:

1) оптимальный план единственный: линия уровня и область допустимых решений Ω в разрешающем положении имеют одну общую точку (рис. 1.4, *а*);

2) оптимальных планов бесконечное множество: в разрешающем положении линия уровня проходит через сторону области допустимых решений (рис. 1.4, *б*);

3) целевая функция не ограничена: линия уровня, сколько бы ее ни перемещали, не может занять разрешающего положения (рис. 1.4, *в*, *г*);

4) область допустимых решений состоит из единственной точки, где целевая функция достигает одновременно и максимального, и минимального значений (рис. 1.4, *д*);

5) задача не имеет решения: область допустимых решений — пустое множество, т. е. система ограничений задачи несовместна (рис. 1.4, *е*).

Пример 1.2. Цех выпускает два вида продукции, используя два вида полуфабрикатов. Продукция используется при комплектовании изделий, при этом на каждую единицу продукции первого вида требуется не более двух единиц продукции второго вида. Нормы расхода a_{ij} полуфабрикатов каждого вида на единицу выпускаемой продукции, общие объемы полуфабрикатов b_i и прибыль c_j от единицы каждой продукции представлены в табл. 1.3. Определить план производства, доставляющий максимум прибыли.

Решение. Пусть $\mathbf{x} = (x_1; x_2)$ — план задачи. Тогда модель задачи запишем в виде $\max Z = 10x_1 + 35x_2$ при ограничениях на полуфабрикаты $x_1 + 2x_2 \leqslant 800$, $6x_1 + 2x_2 \leqslant 2400$, условии комплектности $2x_1 \geqslant x_2$ и неотрицательности переменных $x_1 \geqslant 0$, $x_2 \geqslant 0$.

Таблица 1.3

Полуфабрикаты	Нормы затрат на единицу продукции		Объем полуфабриката
	Π_1	Π_2	
I	1	2	800
II	6	2	2400
Прибыль, ден. ед.	10	35	

Построив соответствующие данным ограничениям-неравенствам граничные прямые $x_1+2x_2 = 800$, $6x_1+2x_2 = 2400$, $2x_1-x_2 = 0$, определим полуплоскости, в которых выполняются эти неравенства (рис. 1.5). Для этого достаточно взять произвольную точку, не лежащую на граничной прямой, и подставить ее координаты в неравенство. Для первых двух неравенств возьмем, например, начало координат $O(0;0)$. Получим истинные утверждения ($0 \leqslant 800$, $0 \leqslant 2400$). Следовательно, первые два неравенства выполняются в полуплоскостях, содержащих точку O. Граничная прямая, соответствующая третьему неравенству, проходит через начало координат. Значит, нужно взять, например, точку (0; 10). Получаем ложное утверждение ($0 \geqslant 10$). Следовательно, третьему неравенству удовлетворяют точки полуплоскости, не содержащей пробной точки (0; 10).

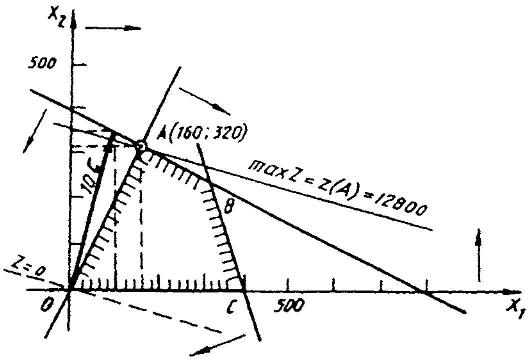


Рис. 1.5

Поскольку $x_1 \geqslant 0$ и $x_2 \geqslant 0$, областью допустимых решений является четырехугольник $OABC$. Далее надо построить вектор $\mathbf{c} = (c_1; c_2)$. Так как он необходим лишь для выяснения направления возрастания целевой функции, иногда для большей наглядности удобно строить вектор $\lambda\mathbf{c}$ ($\lambda > 0$). В нашем примере взято $\lambda = 10$ и построен вектор $\lambda\mathbf{c} = (100; 350)$. Перпендикулярно к этому вектору проводим линию уровня $Z = 0$. Параллельным перемещением прямой $Z = 0$ находим точку A, в которой целевая функция достигает максимума.

Решая совместно уравнения граничных прямых AB и OA, находим координаты точки A: $x_1^* = 160$, $x_2^* = 320$. При этом

$$Z^* = \max Z = z(A) = 12\,800.$$

Итак, по оптимальному плану следует выпускать 160 ед. продукции $П_1$ и 320 ед. продукции $П_2$, что принесет прибыль в 12 800 ден. ед.

Перейдем к геометрической интерпретации ЗЛП с n переменными:

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{1.59}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1,m}), \tag{1.60}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{1.61}$$

Множество планов $\mathbf{x} = (x_1; \dots ; x_n)$, компоненты которых удовлетворяют ограничению-равенству $a_{i1}x_1+\dots+a_{in}x_n = b_i$, геометрически представляет собой гиперплоскость n-мерного пространства. Это выпуклое множество. Множество планов $\mathbf{x} = (x_1; \dots ; x_n)$, компоненты которых удовлетворяют неравенству $a_{i1}x_1 + \dots + a_{in}x_n \leqslant b_i$, образует полупространство n-мерного пространства, которое также является выпуклым множеством.

Множество планов, удовлетворяющих системе ограничений ЗЛП (1.60), (1.61), представляет собой пересечение конечного числа полупространств и потому является выпуклым. Отсюда следует теорема.

Теорема 1.2. *Множество планов ЗЛП выпукло.*

Множество планов ЗЛП в практически важных случаях чаще всего представляет собой либо выпуклый многогранник, либо выпуклую многогранную область.

Целевую функцию (1.59) геометрически можно рассматривать как семейство параллельных гиперплоскостей $c_1x_1 + \dots + c_nx_n = Z$, каждой из которых соответствует определенное значение параметра Z. Вектор $\mathbf{c} = (c_1; \dots ; c_n)$, перпендикулярный к гиперплоскостям $Z = \text{const}$, указывает направление наискорейшего возрастания функции Z.

С учетом сказанного задача (1.59) — (1.61) геометрически сводится к нахождению точки $\mathbf{x}^* = (x_1^*; \dots ; x_n^*)$ многогранника (многоугольной области), определяемого неравенствами (1.60), (1.61), через которую проходит гиперплоскость семейства (1.59), соответствующая наибольшему значению Z.

Графическим методом можно решить ЗЛП с $n > 2$ переменными, если в ее канонической записи число неизвестных n и число линейно независимых уравнений m связаны соотношением $n - m \leqslant 2$. В этом случае каноническую форму задачи преобразовывают, как это описано в § 1.2, в симметричную, которая будет содержать не более двух переменных. Решая эту задачу графически, находят два компонента оптимального плана. Подставляя их в ограничения задачи, определяют и остальные компоненты.

Пример 1.3. Двум погрузчикам разной мощности не более чем за 24 часа нужно погрузить на первой площадке 230 т, на второй 168 т. Первый погрузчик на первой площадке может погрузить 10 т в час, на второй — 12 т. Второй погрузчик на каждой площадке может погрузить по 13 т в час. Стоимость работ, связанных с погрузкой 1 т, первым погрузчиком на первой площадке 8 ден. ед., на второй 7 ден. ед., вторым погрузчиком на первой площадке 12 ден. ед., на второй 13 ден. ед. Нужно составить план работы, т. е. найти, какой объем работ должен выполнить каждый погрузчик на каждой площадке, чтобы стоимость всех работ по погрузке была минимальной. Причем по техническим причинам первый погрузчик на второй площадке должен работать не более 16 часов.

Решение. Обозначим через x_{ij} объем работ (в тоннах) i-го погрузчика ($i = 1, 2$) на j-й площадке ($j = 1, 2$). Условия задачи занесем в табл. 1.4.

Таблица 1.4

j / i	$П_1$	$П_2$	Лимит рабочего времени
I погрузчик	10 8 x_{11}	12 7 x_{12}	24
II погрузчик	13 12 x_{21}	13 13 x_{22}	24
Задание	230	168	

Построим математическую модель задачи. Целевая функция описывает затраты, связанные с выполнением всех работ:

$$8x_{11} + 7x_{12} + 12x_{21} + 13x_{22}.$$

Ограничения на лимиты рабочего времени:

$$\left.\begin{array}{l} x_{11}/10 + x_{12}/12 \leqslant 24, \\ x_{21}/13 + x_{22}/13 \leqslant 24, \end{array}\right\}$$

на необходимость выполнить задание:

$$\left.\begin{array}{l} x_{11} + x_{21} = 230, \\ x_{12} + x_{22} = 168, \end{array}\right\}$$

условие неотрицательности:

$$x_{ij} \geqslant 0 \quad (i, j = 1, 2).$$

Исключим из модели переменные x_{21} и x_{22}. Из ограничений равенств имеем:

$$x_{21} = 230 - x_{11}, \quad x_{22} = 168 - x_{12}.$$

Подставив выражения для x_{21} и x_{22} в ограничения-неравенства и целевую функцию, получим ЗЛП с двумя переменными x_{11} и x_{12}:

$$\min Z = 4944 - 4x_{11} - 6x_{12};$$

$$
\left.\begin{array}{rcrcr}
6x_{11} & + & 5x_{12} & \leqslant & 1440, \\
x_{11} & + & x_{12} & \geqslant & 86, \\
 & & x_{12} & \leqslant & 192, \\
x_{11} & & & \leqslant & 230, \\
 & & x_{12} & \leqslant & 168, \\
\multicolumn{5}{l}{x_{11} \geqslant 0,\ x_{12} \geqslant 0.}
\end{array}\right\}
$$

Очевидно, что целевая функция $Z = 4944 - 4x_{11} - 6x_{12}$ достигает минимального значения при условии, что $Z' = 4x_{11} + 6x_{12}$ принимает максимальное значение. Имеем задачу:

$$\max Z' = 4x_{11} + 6x_{12};$$

$$
\left.\begin{array}{r}
6x_{11} + 5x_{12} \leqslant 1440, \\
x_{11} + x_{12} \geqslant 86, \\
0 \leqslant x_{11} \leqslant 230, \\
0 \leqslant x_{12} \leqslant 168.
\end{array}\right\}
$$

Ее графическое решение представлено на рис. 1.6.

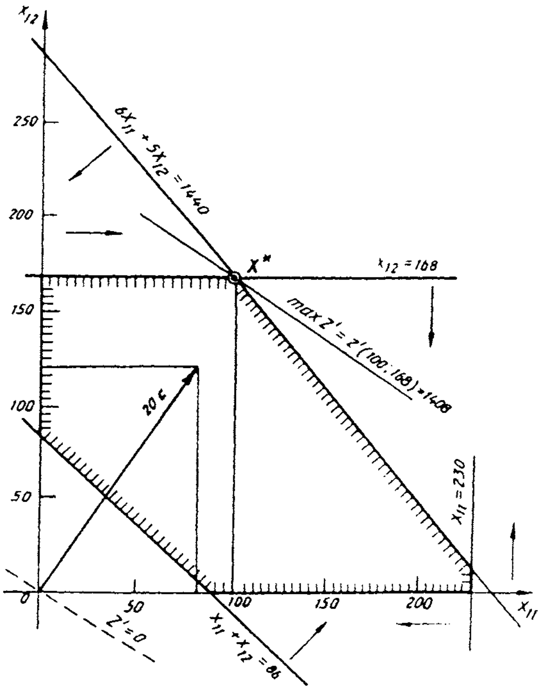


Рис. 1.6

Функция Z' достигает наибольшего значения при $x_{11}^* = 100$, $x_{12}^* = 168$. Из выражений для x_{21} и x_{22} получим: $x_{21}^* = 130$, $x_{22}^* = 0$.

Итак, по оптимальному плану первый погрузчик должен погрузить 100 т на первой площадке и 168 т на второй, второму погрузчику надлежит погрузить 130 т на первой площадке. Стоимость всех работ составит 3536 ден. ед. ($Z^* = 4944 - 4 \cdot 100 - 6 \cdot 168 = 3536$).

Анализируя рис. 1.3 — 1.6, легко заметить, что в ЗЛП с двумя переменными экстремум достигается в вершине области допустимых решений. Оказывается, это не случайный факт. Дальше будет доказано, что и в n-мерном пространстве экстремум целевой функции в ЗЛП достигается в вершине (крайней точке). Напомним, что *крайней (угловой) точкой выпуклого множества* называют точку, которая не может быть представлена выпуклой линейной комбинацией двух других точек этого множества, т. е. точку, не лежащую внутри отрезка, определяемого этими точками.

1.4. СВОЙСТВА РЕШЕНИЙ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Пусть ЗЛП представлена в следующей записи:

$$\max Z = \mathbf{c}\mathbf{x}; \tag{1.62}$$

$$\mathbf{A}_1 x_1 + \mathbf{A}_2 x_2 + \ldots + \mathbf{A}_n x_n = \mathbf{A}_0, \tag{1.63}$$

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ \ldots,\ x_n \geqslant 0. \tag{1.64}$$

Чтобы задача (1.62) — (1.64) имела решение, система ее ограничений (1.63) должна быть совместной. Это возможно, если ранг r системы (число линейно независимых уравнений) не больше числа неизвестных n ($r \leqslant n$). Случай $r > n$ вообще невозможен. При $r = n$ система имеет единственное решение, которое и будет при $x_j \geqslant 0$ ($j = \overline{1,n}$) оптимальным. В этом случае проблема выбора оптимального решения теряет смысл.

Выясним структуру координат угловой точки многогранных решений. Пусть $r < n$. В этом случае система векторов $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$ содержит базис — максимальную линейно независимую подсистему векторов, через которую любой вектор системы может быть выражен как ее линейная комбинация. Базисов, вообще говоря, может быть несколько, но не более C_n^r. Каждый из них состоит точно из r векторов. Переменные ЗЛП, соответствующие r векторам базиса, называют, как известно, *базисными* и обозначают БП. Остальные $n - r$ переменных будут *свободными*, их обозначают СП.

Не ограничивая общности, будем считать, что базис составляют первые m векторов $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$. Этому базису соответствуют базисные переменные $x_1, x_2, \dots, x_m$, а свободными будут переменные $x_{m+1}, x_{m+2}, \dots, x_n$.

Если свободные переменные приравнять нулю, а базисные переменные при этом примут неотрицательные значения, то полученное частное решение системы (1.63) называют *опорным решением* (*планом*).

Теорема 1.3. *Если система векторов* $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$ *содержит* m *линейно независимых векторов* $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$, *то допустимый план*

$$\mathbf{x} = (x_1; x_2; \dots; x_m; \underbrace{0; 0; \dots; 0}_{n-m}) \tag{1.65}$$

является крайней точкой многогранника планов.

Доказательство. Так как векторы $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_m$ линейно независимы, то вектор $\mathbf{A}_0$ может быть выражен через них единственным образом:

$$\mathbf{A}_1 x_1 + \mathbf{A}_2 x_2 + \dots + \mathbf{A}_m x_m = \mathbf{A}_0. \tag{1.66}$$

Предположим, что точка (1.65) не является крайней. Тогда ее можно представить как выпуклую линейную комбинацию двух других различных точек $\mathbf{x}_1$ и $\mathbf{x}_2$ многогранника планов, т. е.

$$\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \quad (\lambda_1 > 0;\ \lambda_2 > 0;\ \lambda_1 + \lambda_2 = 1), \tag{1.67}$$

где

$$\mathbf{x}_1 = (x_1^{(1)}; x_2^{(1)}; \dots; x_m^{(1)}; x_{m+1}^{(1)}; \dots; x_n^{(1)}), \tag{1.68}$$

$$\mathbf{x}_2 = (x_1^{(2)}; x_2^{(2)}; \dots; x_m^{(2)}; x_{m+1}^{(2)}; \dots; x_n^{(2)}). \tag{1.69}$$

Подставив в равенство (1.67) координаты точек (1.65), (1.68) и (1.69), получим:

$$(x_1; x_2; \dots; x_m; 0; \dots; 0) = \lambda_1 (x_1^{(1)}; x_2^{(1)}; \dots; x_m^{(1)};$$

$x_{m+1}^{(1)}; \ldots ; x_n^{(1)}) + \lambda_2(x_1^{(2)}; x_2^{(2)}; \ldots ; x_m^{(2)}; x_{m+1}^{(2)}; \ldots ; x_n^{(2)})$.

Отсюда:

$$\begin{aligned}
x_1 &= \lambda_1 x_1^{(1)} + \lambda_2 x_1^{(2)}, \\
x_2 &= \lambda_1 x_2^{(1)} + \lambda_2 x_2^{(2)}, \\
&\ldots\ldots\ldots\ldots\ldots \\
x_m &= \lambda_1 x_m^{(1)} + \lambda_2 x_m^{(2)}, \\
x_{m+1} &= \lambda_1 x_{m+1}^{(1)} + \lambda_2 x_{m+1}^{(2)} = 0, \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
x_n &= \lambda_1 x_n^{(1)} + \lambda_2 x_n^{(2)} = 0.
\end{aligned}$$

Так как по предположению $\mathbf{x}$ — не крайняя точка, то

$$x_j > 0 \quad (j = \overline{1, m}),\ \lambda_1 > 0,\ \lambda_2.$$

Отсюда следует, что $x_j^{(1)} > 0$, $x_j^{(2)} > 0$ $(j = \overline{1, m})$. Из $x_j = 0$ $(j = \overline{m+1, n})$ и $\lambda_1 > 0$, $\lambda_2 > 0$ следует $x_j^{(1)} = 0$, $x_j^{(2)} = 0$ $(j = \overline{m+1, n})$. Точки $\mathbf{x}_1$ и $\mathbf{x}_2$ имеют ту же структуру, что и точка $\mathbf{x}$, т. е.:

$$\mathbf{x}_1 = (x_1^{(1)}; x_2^{(1)}; \ldots ; x_m^{(1)}; 0; \ldots ; 0), \tag{1.70}$$

$$\mathbf{x}_2 = (x_1^{(2)}; x_2^{(2)}; \ldots ; x_m^{(2)}; 0; \ldots ; 0). \tag{1.71}$$

Поскольку векторы (1.70) и (1.71) — допустимые планы, они должны удовлетворять векторному равенству (1.66). Имеем:

$$\begin{array}{r}
\mathbf{A}_1 x_1^{(1)} + \mathbf{A}_2 x_2^{(1)} + \ldots + \mathbf{A}_m x_m^{(1)} = \mathbf{A}_0 \\
- \\
\mathbf{A}_1 x_1^{(2)} + \mathbf{A}_2 x_2^{(2)} + \ldots + \mathbf{A}_m x_m^{(2)} = \mathbf{A}_0 \\
\hline
\end{array}$$

$$\mathbf{A}_1(x_1^{(1)} - x_1^{(2)}) + \mathbf{A}_2(x_2^{(1)} - x_2^{(2)}) + \ldots + \mathbf{A}_m(x_m^{(1)} - x_m^{(2)}) = \mathbf{A}_0. \tag{1.72}$$

Так как $\mathbf{A}_1, \mathbf{A}_2, \ldots , \mathbf{A}_m$ линейно независимы, то равенство (1.72) выполняется при условии $x_1^{(1)} = x_1^{(2)}$, $x_2^{(1)} = x_2^{(2)}, \ldots$, $x_m^{(1)} = x_m^{(2)}$, т. е. $\mathbf{x}_1 = \mathbf{x}_2$. Пришли к противоречию. Точку $\mathbf{x}$ невозможно представить как выпуклую линейную комбинацию двух различных точек многогранника планов. Точка $\mathbf{x}$ — крайняя точка многогранника решений. $\triangle$

Читателю предлагается доказать утверждение: *если* $\mathbf{x}$ — *крайняя точка многогранника планов, то те векторы* $\mathbf{A}_j$, *которые соответствуют положительным координатам вектора* $\mathbf{x}$, *образуют линейно независимую систему.*

Итак, если допустимый план $\mathbf{x}$ имеет m положительных координат, а все остальные координаты равны нулю, то это *опорный план* ЗЛП. Если положительных координат меньше m, а все остальные равны нулю, то такой план называется *вырожденным опорным планом.* Если же допустимый план имеет больше m положительных координат, то он вообще не соответствует крайней точке многогранника и не является опорным.

Рассмотрим основную теорему линейного программирования. Не сужая область экономических приложений, ЗЛП с неограниченной областью планов формально можно свести к задаче с ограниченной областью введением дополнительного ограничения $x_1 + x_2 + \ldots + x_n \leqslant M$, где M — большое положительное число. Поэтому при рассмотрении теоретических вопросов (для краткости доказательств) в дальнейшем будем предполагать, что множество планов ЗЛП является выпуклым многогранником.

Теорема 1.4. *Если ЗЛП имеет решение, то целевая функция достигает экстремального значения хотя бы в одной из крайних точек многогранника решений. Если же целевая функция достигает экстремального значения более чем в одной крайней точке, то она достигает того же значения в любой точке, являющейся их выпуклой линейной комбинацией.*

Доказательство. Пусть дана ЗЛП: extremum $Z = z(\mathbf{x})$, причем многогранник планов Ω имеет конечное число крайних точек $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. Обозначим через $\mathbf{x}^*$ допустимый план, для которого целевая функция достигает своего, например, максимального значения $\max Z = z(\mathbf{x}^*)$. Отсюда

$$z(\mathbf{x}^*) \geqslant z(\mathbf{x}) \quad (\mathbf{x} \in \Omega). \tag{1.73}$$

Если $\mathbf{x}^*$ совпадает с одной из крайних точек, то первая часть теоремы доказана. Предположим, что $\mathbf{x}^*$ не является крайней точкой многогранника планов. Тогда $\mathbf{x}^*$ можно представить в виде выпуклой линейной комбинации точек $\mathbf{x}_1$,

$\mathbf{x}_2, \dots, \mathbf{x}_k$, т. е.

$$\mathbf{x}^* = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i, \quad \sum_{i=1}^{k} \lambda_i = 1, \ \lambda_i \geqslant 0 \quad (i = \overline{1, k}).$$

В силу линейности $z(\mathbf{x})$ имеем

$$z(\mathbf{x}^*) = \lambda_1 z(\mathbf{x}_1) + \lambda_2 z(\mathbf{x}_2) + \ldots + \lambda_k z(\mathbf{x}_k). \tag{1.74}$$

Обозначим через M максимальное значение функции цели среди всех крайних точек, т. е.

$$M = \max(z(\mathbf{x}_1), z(\mathbf{x}_2), \dots, z(\mathbf{x}_k)). \tag{1.75}$$

Из равенства (1.74) в силу условия (1.75) имеем

$$z(\mathbf{x}^*) \leqslant \lambda_1 M + \lambda_2 M + \ldots + \lambda_k M = M$$

или

$$z(\mathbf{x}^*) \leqslant M. \tag{1.76}$$

Из неравенств (1.73) и (1.76) приходится сделать единственный вывод: $z(\mathbf{x}^*) = M$. Но M — значение целевой функции в одной из крайних точек, поэтому $\mathbf{x}^*$ совпадает с одной из них. Первая часть теоремы доказана.

Для доказательства второй части теоремы допустим, что $z(\mathbf{x})$ достигает максимального значения более чем в одной крайней точке, например в точках $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, т. е.

$$z(\mathbf{x}_1) = z(\mathbf{x}_2) = \ldots = z(\mathbf{x}_p) = M. \tag{1.77}$$

Составим выпуклую линейную комбинацию этих точек:

$$\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \ldots + \lambda_p \mathbf{x}_p, \quad \sum_{i=1}^{p} \lambda_i = 1, \ \lambda_i \geqslant 0 \ (i = \overline{1, p}).$$

Учитывая условие (1.77) и линейность функции Z, получаем

$$z(\mathbf{x}) = \lambda_1 z(\mathbf{x}_1) + \lambda_2 z(\mathbf{x}_2) + \ldots + \lambda_p z(\mathbf{x}_p) =$$
$$= \lambda_1 M + \lambda_2 M + \ldots + \lambda_p M = M,$$

т. е. линейная функция Z принимает максимальное значение в произвольной точке $\mathbf{x}$, являющейся выпуклой линейной комбинацией точек $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$, в которых целевая функция Z принимает максимальное значение. $\triangle$

1.5. СИМПЛЕКСНЫЙ МЕТОД

Общая идея симплексного метода. В связи с основной теоремой линейного программирования естественно возникает мысль о следующем пути решения ЗЛП с любым числом переменных. Найти каким-нибудь способом все крайние точки многогранника планов (их не больше, чем $C_n^r = \frac{n!}{(n-r)!r!}$) и сравнить в них значения целевой функции. Такой путь решения даже с относительно небольшим числом переменных и ограничений практически неосуществим, так как процесс отыскания крайних точек сравним по трудности с решением исходной задачи, к тому же число крайних точек многогранника планов может оказаться весьма большим. В связи с этими трудностями возникла задача рационального перебора крайних точек. Ее суть в следующем. Если известны какая-нибудь крайняя точка и значение в ней целевой функции, то все крайние точки, в которых целевая функция принимает худшее значение, заведомо не нужны. Отсюда естественно стремление найти способ перехода от данной крайней точки к смежной по ребру лучшей, от нее к еще лучшей (не худшей) и т. д. Для этого нужно иметь признак того, что лучших крайних точек, чем данная крайняя точка, вообще нет. В этом и состоит общая идея наиболее широко применяемого в настоящее время *симплексного метода* (*метода последовательного улучшения плана*) для решения ЗЛП. Итак, в алгебраических терминах симплексный метод предполагает: 1) умение находить начальный опорный план; 2) наличие признака оптимальности опорного плана; 3) умение переходить к нехудшему опорному плану.

На рис. 1.7 дана геометрическая интерпретация идеи симплексного метода в случае двух (*а*) и трех (*б*) переменных.

Построение начального опорного плана. Пусть ЗЛП представлена системой ограничений в каноническом виде:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad b_i \geqslant 0 \ (i = \overline{1, m}).$$

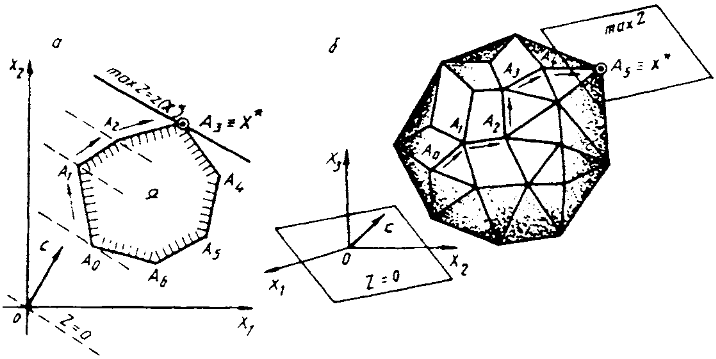


Рис. 1.7

Говорят, что ограничение ЗЛП имеет предпочтительный вид, если при неотрицательности правой части ($b_i \geqslant 0$) левая часть ограничения содержит переменную, входящую с коэффициентом, равным единице, а в остальные ограничения-равенства — с коэффициентом, равным нулю. Например, в системе ограничений

$$\left.\begin{aligned} \underline{x_1} + 2x_2 \qquad\quad - 4x_4 &= 5, \\ 2x_2 + \underline{x_3} + 2x_4 &= 8, \\ x_2 \qquad\quad - 3x_4 &= 3 \end{aligned}\right\}$$

первое и второе ограничения имеют предпочтительный вид, третье — нет.

Если каждое ограничение-равенство ЗЛП в каноническом виде содержит переменную, входящую в левую часть с коэффициентом, равным единице, а во все остальные с коэффициентом, равным нулю (при неотрицательности правых частей), то говорят, что система ограничений представлена в предпочтительном виде. В этом случае легко найти ее опорное решение (базисное с неотрицательными координатами): все свободные переменные нужно приравнять нулю, тогда базисные переменные будут равны свободным членам. Например, в системе ограничений

$$\left.\begin{array}{rl} x_1+\underline{x_2} \qquad\qquad - \ x_5 &= 10, \\ 5x_1 \qquad +\underline{x_3} \qquad +3x_5 &= 80, \\ -5x_1 \qquad\qquad +\underline{x_4}+2x_5 &= 32, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j=\overline{1,5})$$

предпочтительными (базисными) являются переменные x_2, x_3 и x_4, свободными x_1, x_5. Приравняем свободные переменные x_1 и x_5 нулю, тогда базисные переменные примут значения $x_2 = 10$, $x_3 = 80$, $x_4 = 32$. Имеем план $\mathbf{x} = (0;10;80;32;0)$. Если полученный план будет иметь не более m отличных от нуля координат, то, согласно теореме о структуре координат крайней точки, он будет опорным.

Приравнивание предпочтительных переменных к правым частям дает базисное решение, т. е. крайнюю точку многогранника решений. Поэтому предпочтительные переменные — базисные. Переменные, приравниваемые нулю, — свободные.

Пусть, далее, система ограничений имеет вид

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i, \quad b_i \geqslant 0 \quad (i=\overline{1,m}).$$

(Такие ограничения имеют ЗЛП о наилучшем использовании сырья, технологий и т. д.) Сведем задачу к каноническому виду. Для этого добавим к левым частям неравенств дополнительные переменные $x_{n+i} \geqslant 0$ $(i=\overline{1,m})$. Получим, как было показано выше, систему, эквивалентную исходной:

$$\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, \quad b_i \geqslant 0 \quad (i=\overline{1,m}),$$

которая имеет предпочтительный вид. И, следовательно, начальный опорный план примет вид

$$\mathbf{x}_0 = (\underbrace{0;0;\ldots;0}_{n};\underbrace{b_1;b_2;\ldots;b_m}_{m}).$$

В целевую функцию, как отмечалось выше, дополнительные переменные вводятся с коэффициентами, равными нулю:

$$c_{n+i} = 0 \quad (i=\overline{1,m}).$$

Пусть, далее, система ограничений имеет вид

$$\sum_{j=1}^{n} a_{ij}x_j \geqslant b_i, \quad b_i \geqslant 0 \quad (i = \overline{1,m}).$$

Сведем ее к эквивалентной вычитанием дополнительных переменных $x_{n+i} \geqslant 0 \quad (i = \overline{1,m})$ из левых частей неравенств системы. Получим систему

$$\sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i, \quad b_i \geqslant 0 \quad (i = \overline{1,m}).$$

Однако теперь система ограничений не имеет предпочтительного вида, так как дополнительные переменные x_{n+i} входят в левую часть (при $b_i \geqslant 0$) с коэффициентами, равными -1. Поэтому, вообще говоря, базисный план

$$\mathbf{x}_0 = (\underbrace{0; 0; \ldots; 0;}_{n} -b_1; -b_2; \ldots; -b_m)$$

является недопустимым. В этом случае вводится так называемый *искусственный базис.* К левым частям ограничений-равенств, не имеющих предпочтительного вида, добавляют искусственные переменные w_i. В целевую функцию переменные w_i вводят с коэффициентом M в случае решения задачи на минимум и с коэффициентом $-M$ для задачи на максимум, где M — большое положительное число. Полученная задача называется *M-задачей,* соответствующей исходной. Она всегда имеет предпочтительный вид.

Пусть исходная ЗЛП имеет вид:

$$\max(\min) Z = \sum_{j=1}^{n} c_j x_j; \tag{1.78}$$

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad b_i \geqslant 0 \quad (i = \overline{1,m}), \tag{1.79}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}), \tag{1.80}$$

причем ни одно из ограничений не имеет предпочтительной переменной. M-задача запишется так:

$$\max(\min)\,\overline{Z} = \sum_{j=1}^{n} c_j x_j - (+) \sum_{i=1}^{m} M w_i; \qquad (1.81)$$

$$\sum_{j=1}^{n} a_{ij} x_j + w_i = b_i \quad (i = \overline{1,m}), \qquad (1.82)$$

$$x_j \geqslant 0 \ (j = \overline{1,n}), \quad w_i \geqslant 0 \quad (i = \overline{1,m}), \qquad (1.83)$$

где знак "$-$" в функции (1.81) относится к задаче на максимум. Задача (1.81) — (1.83) имеет предпочтительный вид. Ее начальный опорный план

$$\mathbf{x}_0 = (\underbrace{0; 0; \ldots; 0;}_{n} b_1; b_2; \ldots; b_m).$$

Если некоторые из уравнений (1.79) имеют предпочтительный вид, то в них не следует вводить искусственные переменные.

Теорема 1.5. *Если в оптимальном плане*

$$\overline{\mathbf{x}} = (x_1; x_2; \ldots; x_n; w_1; w_2; \ldots; w_m) \qquad (1.84)$$

M-задачи (1.81) — (1.83) все искусственные переменные $w_i = 0 \ (i = \overline{1,m})$, то план

$$\mathbf{x} = (x_1; x_2; \ldots; x_n) \qquad (1.85)$$

является оптимальным планом исходной задачи (1.78) — (1.80).

Доказательство. Если план (1.84) является оптимальным для M-задачи, то план (1.85) будет допустим для исходной задачи (1.78) — (1.80), причем

$$\overline{z}(\overline{\mathbf{x}}) = z(\mathbf{x}). \qquad (1.86)$$

Это следует из того, что план $\overline{\mathbf{x}}$ отличается от плана $\mathbf{x}$ только последними компонентами, равными нулю. Следовательно, если в задачу (1.81) — (1.83) подставить $w_i = 0$ $(i = \overline{1, m})$, то получим задачу (1.78) — (1,80). Докажем, что $\mathbf{x}$ — оптимальный план исходной задачи. Допустим, что $\mathbf{x}$ не является оптимальным планом исходной задачи. Тогда существует план $\mathbf{x}^*$, для которого в случае задачи на максимум

$$z(\mathbf{x}^*) \geqslant z(\mathbf{x}). \tag{1.87}$$

В силу условий (1.86) и (1.87) $\bar{z}(\overline{\mathbf{x}}^*) = z(\mathbf{x}^*) \leqslant z(\mathbf{x}) = \bar{z}(\overline{\mathbf{x}})$, т. е.

$$\bar{z}(\overline{\mathbf{x}}^*) \geqslant \bar{z}(\overline{\mathbf{x}}), \tag{1.88}$$

где $\overline{\mathbf{x}}^*$ — план M-задачи, соответствующий плану $\mathbf{x}^*$ исходной задачи.

Из неравенства (1.88) следует, что план $\overline{\mathbf{x}}$ не является оптимальным для M-задачи. Это, однако, противоречит условию теоремы. Противоречие вызвано неверным предположением о неоптимальности плана $\overline{\mathbf{x}}$. $\triangle$

Аналогично доказывается теорема для задачи на минимум.

Для того чтобы решить задачу с ограничениями, не имеющими предпочтительного вида, вводят искусственный базис и решают расширенную M-задачу, которая имеет начальный опорный план

$$\overline{\mathbf{x}}_0 = (\underbrace{0; 0; \ldots; 0;}_{n}\ b_1; b_2; \ldots; b_m).$$

Решение исходной задачи симплексным методом путем введения искусственных переменных w_i называется *симплексным методом с искусственным базисом.*

Если в результате применения симплексного метода к расширенной задаче получен оптимальный план, в котором все искусственные переменные $w_i^* = 0$, то его первые n компоненты дают оптимальный план исходной задачи.

Теорема 1.6. *Если в оптимальном плане M-задачи хотя бы одна из искусственных переменных отлична от нуля, то исходная задача не имеет допустимых планов, т. е. ее условия несовместны.*

Доказательство. Рассмотрим, например, задачу максимизации. Пусть $w_{i_0}^* > 0$, тогда целевая функция (1.81) примет вид

$$\overline{Z} = \sum_{j=1}^{n} c_j x_j - M w_{i_0}.$$

Поскольку $M > 0$ — достаточно большое число, то при $w_{i_0} \to \to \infty$ $\quad Z \to -\infty$, что противоречит максимизации целевой функции. Δ

Признак оптимальности опорного плана. Симплексные таблицы. Любую ЗЛП, как было показано выше, можно представить в эквивалентном предпочтительном виде:

$$\max(\min) Z = \sum_{j=1}^{n} c_j x_j; \tag{1.89}$$

$$x_i + \sum_{j=m+1}^{n} \alpha_{ij} x_j = \beta_i, \quad \beta_i \geqslant 0 \quad (i = \overline{1,m}), \tag{1.90}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{1.91}$$

Выразим базисные переменные $x_1, x_2, \ldots, x_m$ из равенств (1.90) через свободные $x_{m+1}, x_{m+2}, \ldots, x_n$ и подставим их в целевую функцию. После группировки подобных членов получим

$$Z = (c_1\beta_1 + c_2\beta_2 + \ldots + c_m\beta_m) - ((c_1\alpha_{1,m+1} +$$
$$+ c_2\alpha_{2,m+1} + \ldots + c_m\alpha_{m,m+1}) - c_{m+1})x_{m+1} + ((c_1\alpha_{1,m+2} +$$
$$+ c_2\alpha_{2,m+2} + \ldots + c_m\alpha_{m,m+2}) - c_{m+2})x_{m+2} + \ldots + ((c_1\alpha_{1n} +$$
$$+ c_2\alpha_{2n} + \ldots + c_m\alpha_{mn}) - c_n)x_n. \tag{1.92}$$

Введем обозначения:

$$\Delta_0 = c_1\beta_1 + c_2\beta_2 + \ldots + c_m\beta_m = \mathbf{c}_\text{Б}\mathbf{A}_0, \tag{1.93}$$

$$\Delta_j = (c_1\alpha_{1j} + c_2\alpha_{2j} + \ldots + c_m\alpha_{mj}) - c_j = \mathbf{c}_\text{Б}\mathbf{A}_j - c_j \quad (j = \overline{1,n}), \tag{1.94}$$

где $\mathbf{c}_\text{Б} = (c_1; c_2; \ldots; c_m)$ — вектор коэффициентов целевой функции при базисных переменных; $\mathbf{A}_0 = (\beta_1; \beta_2; \ldots; \beta_m)^T$ —

вектор-столбец свободных членов; $\mathbf{A}_j = (\alpha_{1j}; \alpha_{2j}; \ldots; \alpha_{mj})^T$ — вектор-столбец коэффициентов при переменных x_j.

С учетом равенств (1.92) — (1.94) задача (1.89) — (1.91) примет вид:

$$\max(\min) Z = \Delta_0 - \sum_{j=m+1}^{n} \Delta_j x_j; \tag{1.95}$$

$$x_i + \sum_{j=m+1}^{n} \alpha_{ij} x_j = \beta_i \quad (i = \overline{1,m}), \tag{1.96}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}), \tag{1.97}$$

где $\Delta_0 = \mathbf{c}_{\text{Б}}\mathbf{A}_0$; $\Delta_j = \mathbf{c}_{\text{Б}}\mathbf{A}_j - c_j \quad (j = \overline{1,n})$.

Задачу (1.95) — (1.97) записывают в таблицу, которая называется *симплексной* (табл. 1.5). Последнюю, $(m+1)$-ю, строку называют *индексной строкой* (*строкой целевой функции*), число $\Delta_0 = \mathbf{c}_{\text{Б}}\mathbf{A}_0$ — значение целевой функции для начального опорного плана $\mathbf{x}_0$, т.е. $\Delta_0 = z(\mathbf{x}_0) = \mathbf{c}_{\text{Б}}\mathbf{A}_0$. Числа $\Delta_j = \mathbf{c}_{\text{Б}}\mathbf{A}_j - c_j \quad (j = \overline{1,n})$ называют *оценками свободных переменных.*

Таблица 1.5

БП	$\mathbf{c}_{\text{Б}}$	$\mathbf{A}_0$	x_1	x_2	...	x_i	...	x_m	x_{m+1}	...	x_j	...	x_n
			c_1	c_2	...	c_i	...	c_m	c_{m+1}	...	c_j	...	c_n
x_1	c_1	β_1	1	0	...	0	...	0	$\alpha_{1,m+1}$	...	α_{1j}	...	α_{1n}
x_2	c_2	β_2	0	1	...	0	...	0	$\alpha_{2,m+1}$	...	α_{2j}	...	α_{2n}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
x_i	c_i	β_i	0	0	...	1	...	0	$\alpha_{i,m+1}$	...	α_{ij}	...	α_{in}
...	...	...	...	...	...	...	...	...	...	...	...	...	...
x_m	c_m	β_m	0	0	...	0	...	1	$\alpha_{m,m+1}$	...	α_{mj}	...	α_{mn}
$z_j - c_j$		Δ_0	0	0	...	0	...	0	Δ_{m+1}	...	Δ_j	...	Δ_n

Теорема 1.7. *Пусть исходная задача решается на максимум. Если для некоторого опорного плана все оценки Δ_j $(j = \overline{1,n})$ неотрицательны, то такой план оптимален.*

Доказательство. Так как $Z = \Delta_0 - \sum\limits_{j=m+1}^{n} \Delta_j x_j$ и по условию $\Delta_j \quad (j = \overline{1,n})$, то Z достигает максимального значения при $\sum\limits_{j=m+1}^{n} \Delta_j x_j = 0$. Это возможно лишь при $x_{m+1} = 0$, $x_{m+2} = 0, \ldots, x_n = 0$, т. е. опорный план

$$(\beta_1; \beta_2; \ldots; \beta_m; \underbrace{0; \ldots; 0}_{n-m})$$

оптимален. $\triangle$

Теорема 1.8. *Если исходная задача решается на минимум и для некоторого опорного плана все оценки Δ_j $(j = \overline{1, n})$ неположительны, то такой план оптимален.*

Доказательство аналогично предыдущему случаю.

Пример 1.4. Решить ЗЛП:

$$\min Z = 2x_1 - x_2 + 3x_3 - 2x_4 + x_5;$$

$$\left.\begin{aligned} x_2 + 0,5x_3 \quad\quad + 0,5x_5 &= 1,5, \\ x_3 + x_4 \quad\quad &= 2, \\ x_1 \quad - 0,5x_3 \quad\quad + 0,5x_5 &= 0,5, \end{aligned}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1, 5}).$$

Решение. Система ограничений задачи имеет предпочтительный вид, так как каждое уравнение-ограничение содержит переменную с коэффициентом, равным единице, которая во все остальные уравнения входит с коэффициентом, равным нулю. Это переменные x_2, x_4, x_1. Они и составят базис. Заносим условие задачи в симплексную таблицу (табл. 1.6). Таблица содержит $n + 3$ столбцов, где n — число переменных в предпочтительном виде, и $m + 2$ строк, где m — число ограничений-равенств.

Таблица 1.6

БП	$\mathbf{c}_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5
			2	−1	3	−2	1
x_2	−1	1,5	0	1	0,5	0	0,5
x_4	−2	2	0	0	1	1	0
x_1	2	0,5	1	0	− 0,5	0	0,5
$z_j - c_j$		− 4,5	0	0	− 6,5	0	− 0,5

В столбце БП записываются базисные переменные. Столбец $\mathbf{c}_Б$ содержит коэффициенты целевой функции, стоящие при базисных переменных. Для нашего случая $c_2 = -1$, $c_4 = -2$ и $c_1 = 2$. Столбец $\mathbf{A}_0$ —

столбец свободных членов β_i системы ограничений. Основное поле таблицы занимают коэффициенты α_{ij} системы ограничений. Остановимся подробнее на заполнении индексной строки $z_j - c_j$. Здесь расположены значение функции цели для начального опорного плана $\mathbf{x}_0$, т. е. $z(\mathbf{x}_0) = \Delta_0 = \mathbf{c}_Б \mathbf{A}_0$, и оценки индексной строки $\Delta_j = \mathbf{c}_Б \mathbf{A}_j - c_j$:

$$\begin{aligned}
\Delta_0 &= (-1)1,5 + (-2)2 + 2 \cdot 0,5 = -4,5,\\
\Delta_1 &= (-1)0 + (-2)0 + 2 \cdot 1 - 2 = 0,\\
\Delta_2 &= (-1)1 + (-2)0 + 2 \cdot 0 - (-1) = 0,\\
\Delta_3 &= (-1)0,5 + (-2)1 + 2(-0,5) - 3 = -6,5,\\
\Delta_4 &= (-1)0 + (-2)1 + 2 \cdot 0 - (-2) = 0,\\
\Delta_5 &= (-1)0,5 + (-2)0 + 2 \cdot 0,5 - 1 = -0,5.
\end{aligned}$$

Начальный опорный план задачи:

$$\mathbf{x}_0 = (0,5;\ 1,5;\ 0;\ 2;\ 0), \quad z(\mathbf{x}_0) = -4,5.$$

Так как все оценки индексной строки Δ_j $(j = \overline{1,5})$ неположительны, то план $\mathbf{x}_0$ оптимален:

$$\mathbf{x}^* = (0,5;\ 1,5;\ 0;\ 2;\ 0), \quad z(\mathbf{x}^*) = -4,5.$$

Переход к нехудшему опорному плану. Пусть решается ЗЛП с системой ограничений в предпочтительном виде

$$x_i + \sum_{j=m+1}^{n} \alpha_{ij} x_j = \beta_i, \quad \beta_i \geqslant 0 \ (i = \overline{1,m}). \tag{1.98}$$

Ее начальный опорный план $\mathbf{x}_0 = (\beta_1; \ldots; \beta_m; 0; \ldots; 0)$. Значение целевой функции $z(\mathbf{x}_0) = \mathbf{c}_Б \mathbf{A}_0 = \Delta_0$.

Рассмотрим задачу на максимум. Если все $\Delta_j \geqslant 0$, то опорный план $\mathbf{x}_0$ оптимален. Пусть существует j_0, для которого $\Delta_{j_0} < 0$. Вектор-столбец $\mathbf{A}_{j_0}$, для которого $\Delta_{j_0} < 0$, называется *разрешающим*, соответствующая переменная x_{j_0} — *перспективной*. Попытаемся, не изменяя нулевых значений свободных переменных $x_{m+1}, x_{m+2}, \ldots, x_n$, кроме x_{j_0}, увеличить значение целевой функции Z за счет увеличения переменной $x_{j_0} > 0$. Однако увеличивать x_{j_0} следует осторожно, так как выбор x_{j_0} влияет на значения $x_1, \ldots, x_m$, которые должны быть неотрицательными. Из системы ограничений (1.98) в силу того, что

$$x_{m+1} = 0, \ \ldots, \ x_{j_0-1} = 0, \ x_{j_0} > 0, \ x_{j_0+1} = 0, \ \ldots, \ x_n = 0,$$

имеем

$$x_i = \beta_i - \alpha_{ij_0} x_{j_0} \quad (i = \overline{1,m}).$$

При значительном увеличении x_{j_0} может случиться, что для некоторого i соответствующее β_i станет меньше $\alpha_{ij_0}x_{j_0}$, и, следовательно, получим $x_i < 0$, что недопустимо. В случае, если $\alpha_{ij_0} \leqslant 0$ $(i = \overline{1, m})$, такого нарушения не произойдет.

Итак, не ограничивая общности, будем считать, что, например, первые k коэффициентов $\alpha_{ij_0} > 0$ $(i = \overline{1, k};\ k < m)$. Тогда x_{j_0} можно увеличивать до тех пор, пока $\beta_i - \alpha_{ij_0}x_{j_0} \geqslant 0$ $(i = \overline{1, k})$, откуда имеем:

$$x_{j_0} \leqslant \beta_i/\alpha_{ij_0}, \quad \alpha_{ij_0} > 0 \quad (i = \overline{1, k}).$$

Найдем среди отношений β_i/α_{ij_0} $(i = \overline{1, k})$ наименьшее. Назовем его *наименьшим симплексным отношением* и обозначим буквой θ, т. е.

$$x_{j_0} = \min\left\{\frac{\beta_i}{\alpha_{ij_0}}\right\} = \frac{\beta_{i_0}}{\alpha_{i_0j_0}} = \theta.$$

Если это условие выполняется при нескольких i, то в качестве i_0 можно выбрать любое. Строку i_0 называют *разрешающей*, элемент же $\alpha_{i_0j_0}$ — *разрешающим* (или *ключевым*). Переменная x_{i_0}, присутствующая в базисе, является неперспективной, ее следует вывести из базиса.

Полагая $x_{m+1} = 0, \ldots, x_{j_0-1} = 0$, $x_{j_0} = \theta$, $x_{j_0+1} = 0, \ldots,$ $x_n = 0$, из равенств (1.98) находим:

$$x_1 = \beta_1 - \alpha_{1j_0}\theta, \ \ldots, \ x_{i_0-1} = \beta_{i_0-1} - \alpha_{i_0-1,j_0}\theta, \ x_{i_0} = 0,$$

$$x_{i_0+1} = \beta_{i_0+1} - \alpha_{i_0+1,j_0}\theta, \ \ldots, \ x_m = \beta_m - \alpha_{mj_0}\theta.$$

Новый базис будет состоять из переменных $x_1, x_2, \ldots,$ $x_{i_0-1}, x_{j_0}, x_{i_0+1}, \ldots, x_n$, а соответствующий ему опорный план примет вид

$$\mathbf{x}_1 = (\beta_1 - \alpha_{1j_0}\theta, \ \beta_2 - \alpha_{2j_0}\theta, \ \ldots, \ \beta_{i_0-1} - \alpha_{i_0-1,j_0}\theta, \ \underbrace{0,}_{i_0}$$

$$\beta_{i_0+1} - \alpha_{i_0+1,j_0}\theta, \ \ldots, \ \beta_m - \alpha_{mj_0}\theta, 0, \ldots, 0, \underbrace{\theta,}_{j_0} \ 0, \ldots, 0).$$

В результате преобразований получен новый опорный план $\mathbf{x}_1$, в котором переменная x_{i_0} заменена на x_{j_0}, причем

$\Delta_0 - \Delta_{j_0}\theta = z(\mathbf{x}_0) - \Delta_{j_0}\theta$. Но $\Delta_{j_0} < 0$, следовательно, $z(\mathbf{x}_1) \geqslant$ $\geqslant z(\mathbf{x}_0)$. Новый план $\mathbf{x}_1$ не хуже начального $\mathbf{x}_0$.

Практика показывает, что в случае решения задачи на максимум число шагов, как правило, уменьшается, если разрешающий столбец выбрать по правилу $\max |\Delta_j|$ $(\Delta_j < 0)$, т. е. в базис вводить переменную, соответствующую максимальной по абсолютной величине отрицательной оценке.

В случае задачи на минимум разрешающий столбец нужно выбирать по правилу

$$\max \Delta_j \quad (\Delta_j > 0).$$

Далее процесс повторяется. Проверяем, является ли план $\mathbf{x}_1$ оптимальным. Если да, то задача решена. Если нет, то переходим к нехудшему опорному плану $\mathbf{x}_2$, смежному с $\mathbf{x}_1$, и т. д.

Симплексные преобразования. Чтобы завершить шаг преобразований, ведущих к новому опорному плану в новом базисе, выразим новую базисную переменную x_{j_0} из уравнения с номером $i = i_0$ системы (1.96) через свободные переменные $x_{m+1}, \ldots, x_{j_0-1}, x_{i_0}, x_{j_0+1}, \ldots, x_n$:

$$x_{j_0} = \frac{\beta_{i_0}}{\alpha_{i_0j_0}} - \left(\frac{\alpha_{i_0,m+1}}{\alpha_{i_0j_0}}x_{m+1} + \ldots + \frac{\alpha_{i_0,j_0-1}}{\alpha_{i_0j_0}}x_{j_0-1} + \right.$$
$$\left. + \frac{1}{\alpha_{i_0j_0}}x_{i_0} + \frac{\alpha_{i_0,j_0+1}}{\alpha_{i_0j_0}}x_{j_0+1} + \ldots + \frac{\alpha_{i_0n}}{\alpha_{i_0j_0}}x_n\right). \tag{1.99}$$

Подставим выражение (1.99) в остальные ограничения системы (1.96):

$$x_i = \frac{\beta_i\alpha_{i_0j_0} - \beta_{i_0}\alpha_{ij_0}}{\alpha_{i_0j_0}} - \left(\frac{\alpha_{i,m+1}\alpha_{i_0j_0} - \alpha_{i_0,m+1}\alpha_{ij_0}}{\alpha_{i_0j_0}}x_{m+1} + \ldots + \right.$$
$$\left. + \frac{\alpha_{ij_0}}{\alpha_{i_0j_0}}x_{i_0} + \ldots + \frac{\alpha_{in}\alpha_{i_0j_0} - \alpha_{i_0n}\alpha_{ij_0}}{\alpha_{i_0j_0}}x_n\right)$$
$$(i = 1, 2, \ldots, i-1, i+1, \ldots, m). \tag{1.100}$$

Аналогично, подставив выражение x_{j_0} из равенства (1.99) в целевую функцию (1.95), получим

$$Z = \frac{\Delta_0\alpha_{i_0j_0} - \Delta_{j_0}\beta_{i_0}}{\alpha_{i_0j_0}} - \left(\frac{\Delta_{m+1}\alpha_{i_0j_0} - \Delta_{j_0}\alpha_{i,m+1}}{\alpha_{i_0j_0}}x_{m+1} + \ldots + \right.$$

$$+\frac{\Delta_{j_0}}{\alpha_{i_0 j_0}} x_{i_0} + \ldots + \frac{\Delta_n \alpha_{i_0 j_0} - \Delta_{j_0} \alpha_{i_0 n}}{\alpha_{i_0 j_0}} x_n \Big). \qquad (1.101)$$

Преобразование ЗЛП к новому базису назовем *симплексным преобразованием.* Из равенств (1.99) — (1.101) вытекают правила для перехода к следующей симплексной таблице.

1. Из выражения (1.99) следует, что элементы строки i_0 новой таблицы равны соответствующим элементам разрешающей строки старой таблицы, деленным на разрешающий элемент:

$$\beta'_{i_0} = \frac{\beta_{i_0}}{\alpha_{i_0 j_0}}, \quad \alpha'_{i_0 j} = \frac{\alpha_{i_0 j}}{\alpha_{i_0 j_0}} \quad (j = \overline{1, n}).$$

2. Элементы разрешающего столбца j_0 новой таблицы равны нулю, за исключением $\alpha'_{i_0 j_0} = 1$:

$$\alpha'_{i j_0} = 0 \quad (i \neq i_0), \quad \alpha'_{i_0 j_0} = 1.$$

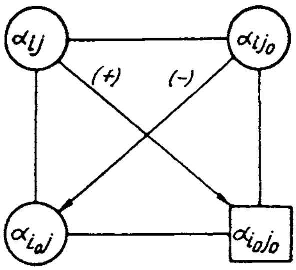


Рис. 1.8

3. Чтобы найти любой другой элемент новой симплексной таблицы, нужно воспользоваться правилом прямоугольника (рис. 1.8), вытекающим из формул:

$$\left.\begin{aligned} \beta'_i &= \frac{\beta_i \alpha_{i_0 j_0} - \beta_{i_0} \alpha_{i j_0}}{\alpha_{i_0 j_0}} = \beta_i - \frac{\beta_{i_0} \alpha_{i j_0}}{\alpha_{i_0 j_0}}, \\ \alpha'_{ij} &= \frac{\alpha_{ij} \alpha_{i_0 j_0} - \alpha_{i_0 j} \alpha_{i j_0}}{\alpha_{i_0 j_0}} = \alpha_{ij} - \frac{\alpha_{i_0 j} \alpha_{i j_0}}{\alpha_{i_0 j_0}}, \\ &(i \neq i_0;\ j \neq j_0). \end{aligned}\right\} \qquad (1.102)$$

Для этого в исходной таблице выделяют прямоугольник, вершинами которого служат нужные для вычисления элементы (см. рис. 1.8). Диагональ, содержащую разрешающий и искомый элементы новой таблицы, называют *главной*, а другую — *побочной. Чтобы получить элемент* α'_{ij}, $(i = i_0;\ j \neq j_0)$ *новой симплексной таблицы, нужно из произведения угловых элементов главной диагонали вычесть произведение угловых элементов побочной диагонали и полученное число разделить на разрешающий элемент, выделенный рамкой.* Это *правило прямоугольника.*

Как следует из равенств (1.102), любой элемент новой таблицы можно найти по *правилу треугольника: для получения любого элемента новой симплексной таблицы нужно из соответствующего элемента прежней таблицы вычесть произведение элемента разрешающей строки на элемент разрешающего столбца, разделенное на разрешающий элемент.*

4. Как следует из равенства (1.101), по этим же правилам могут быть вычислены все элементы индексной строки Δ'_j $(j = \overline{1, n})$ и новое значение целевой функции Δ'_0:

$$\left.\begin{aligned} \Delta'_j &= \frac{\Delta_j \alpha_{i_0 j_0} - \Delta_{j_0} \alpha_{i_0 j}}{\alpha_{i_0 j_0}} = \Delta_j - \frac{\Delta_{j_0} \alpha_{i_0 j}}{\alpha_{i_0 j_0}}, \\ \Delta'_0 &= \frac{\Delta_0 \alpha_{i_0 j_0} - \Delta_{j_0} \beta_{i_0}}{\alpha_{i_0 j_0}} = \Delta_0 - \frac{\Delta_{j_0} \beta_{i_0}}{\alpha_{i_0 j_0}}. \end{aligned}\right\} \tag{1.103}$$

Шаг симплексного метода, позволяющий перейти от одного опорного плана к другому нехудшему, называется *итерацией.* Таким образом, симплексный метод является итеративным методом последовательного улучшения плана. Из второго соотношения (1.103) следует, что значения целевой функции на двух последовательных итерациях связаны равенством

$$\Delta_0^{(k+1)} = \Delta_0^{(k)} - \Delta_{j_0}^{(k)} \theta_{i_0 j_0}^{(k)},$$

т. е.

$$z(\mathbf{x}_{k+1}) = z(\mathbf{x}_k) - \Delta_{j_0}^{(k)} \theta_{i_0 j_0}^{(k)}, \tag{1.104}$$

где $z(\mathbf{x}_k)$, $z(\mathbf{x}_{k+1})$ — значения функции цели на k-й и $(k+1)$-й итерациях; $\Delta_{j_0}^{(k)}$ — оценка свободной переменной x_{j_0}, вводимой в базис; $\theta_{i_0 j_0}^{(k)}$ — симплексное отношение k-й итерации.

В результате ввода свободной переменной x_{j_0} в базис изменение целевой функции Z, как это следует из равенства

(1.104), составит

$$\Delta Z^{(k)} = z(\mathbf{x}_{k+1}) - z(\mathbf{x}_k) = -\Delta_{j_0}^{(k)}\theta_{i_0 j_0}^{(k)}.$$

В очередном опорном плане новая базисная переменная x_{j_0} примет значение, равное $\theta_{i_0 j_0}^{(k)}$. Ее прежнее значение (как свободной переменной) равно нулю. Поэтому изменение Δx_{j_0} переменной x_{j_0} в результате введения ее в новый базис равно $\theta_{i_0 j_0}^{(k)}$. В таком случае $\Delta Z^{(k)}$ можно переписать в виде $\Delta Z^{(k)} = -\Delta_{j_0}^{(k)}\Delta x_{j_0}$, откуда

$$\Delta_{j_0}^{(k)} = -\Delta Z^{(k)}/\Delta x_{j_0}.$$

Если, в частности, $\Delta x_{j_0} = 1$, то $\Delta_{j_0}^{(k)} = -\Delta Z^{(k)}$.

Таким образом, оценка $\Delta_{j_0}^{(k)}$ свободной переменной x_{j_0} характеризует изменение целевой функции Z, приходящееся на каждую единицу изменения значения переменной x_{j_0}. Этим и объясняется термин "оценка".

Контроль вычислений. Коэффициенты целевой функции, как показано выше, могут быть рассчитаны по формулам (1.93), (1.94), которые применяют для контроля вычислений после каждого шага симплексных преобразований. Если же их используют для заполнения индексной строки, то для контроля вычислений следует применять формулы (1.103).

Пример 1.5. Решить симплекс-методом ЗЛП:

$$\max Z = 14x_1 - 5x_2 + 2x_3 - x_4 + 8x_5;$$

$$\left.\begin{array}{rl} x_1 + \underline{x_2} \qquad\qquad - x_5 &= 5, \\ 5x_1 \qquad + \underline{x_3} \qquad + 3x_5 &= 41, \\ -5x_1 \qquad\qquad + \underline{x_4} + 4x_5 &= 15, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,5}).$$

Решение. Так как задача имеет предпочтительный вид, занесем ее условия в симплексную таблицу 1.7 (итерация 0). Начальный опорный план $\mathbf{x}_0 = (0; 5; 41; 15; 0)$, $z(\mathbf{x}_0) = 42$.

Для задачи максимизации условием оптимальности опорного плана является неотрицательность оценок. В данном случае две оценки отрицательны. Наибольшая из них по абсолютной величине соответствует столбцу переменной x_1. Этот столбец и назначим разрешающим. Для определения разрешающей строки находим минимальное симплексное отношение:

$$\min_{a_{ij_0}>0}\left\{\frac{b_i}{a_{ij_0}}\right\} = \min_{a_{i1}>0}\left\{\frac{b_i}{a_{i1}}\right\} = \min\left\{\frac{5}{1}, \frac{41}{5}\right\} = 5, \quad i_0 = 1.$$

Таблица 1.7

Номер итерации	БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	Симплексные отношения
				14	−5	2	−1	8	
0	x_2	− 5	5	$\boxed{1}$	1	0	0	− 1	$5/1 = 5$
	x_3	2	41	5	0	1	0	3	$41/5 = 8,2$
	x_4	− 1	15	− 5	0	0	1	4	—
	$z_j - c_j$		42	− 4	0	0	0	− 1	—
1	x_1	14	5	1	1	0	0	− 1	—
	x_3	2	16	0	− 5	1	0	$\boxed{8}$	$16/8 = 2$
	x_4	− 1	40	0	5	0	1	− 1	—
	$z_j - c_j$		62	0	4	0	0	− 5	—
2	x_1	14	7	1	3/8	1/8	0	0	—
	x_5	8	2	0	− 5/8	1/8	0	1	—
	x_4	− 1	42	0	35/8	1/8	1	0	—
	$z_j - c_j$		72	0	7/8	5/8	0	0	—

Оно соответствует первой строке, которая и будет разрешающей.

Следовательно, элемент $a_{11} = 1$ — разрешающий. В итерации 0 табл. 1.7 он выделен рамкой. Переменную x_2 выведем из базиса, а x_1 введем в базис. Разрешающую строку делим на разрешающий элемент. Элементы разрешающего столбца заполняем нулями, кроме $a'_{11} = 1$, а все остальные элементы таблицы пересчитываем по правилу треугольника (прямоугольника).

Например, $a'_{25} = 3 - \frac{5(-1)}{1} = 8$ и т.д. (итерация 1). По этому же правилу заполняются оценки индексной строки, например (итерация 1):

$$\Delta_0 = 42 - \frac{5(-4)}{1} = 62, \quad \Delta_5 = (-1) - \frac{(-4)(-1)}{1} = -5, \ldots$$

Заполнив таблицу, после каждой итерации делаем контрольные проверки, используя соответствующие формулы.

Для итерации 1:

$$\begin{aligned}
\Delta'_0 &= 14 \cdot 5 + 2 \cdot 16 + (-1) \cdot 40 = 62,\\
\Delta'_1 &= 14 \cdot 1 + 2 \cdot 0 + (-1) \cdot 0 - 14 = 0,\\
\Delta'_2 &= 14 \cdot 1 + 2(-5) + (-1) \cdot 5 = 4,\\
\Delta'_3 &= 14 \cdot 0 + 2 \cdot 1 + (-1) \cdot 0 - 2 = 0,\\
\Delta'_4 &= 14 \cdot 0 + 2 \cdot 0 + (-1) \cdot 1 - (-1) = 0,\\
\Delta'_5 &= 14(-1) + 2 \cdot 8 + (-1)(-1) - 8 = -5,
\end{aligned}$$

$$z(\mathbf{x}_1) = 62, \quad z(\mathbf{x}_0) = 42, \quad \Delta_1^{(0)} = 4, \quad \theta_{1,1}^{(0)} = 5, \quad 62 = 42 - (-4)5.$$

Так как существует отрицательная оценка $\Delta_5 = -5$, опорный план $\mathbf{x}_1 = (5; 0; 16; 40; 0)$ неоптимален. Введем в базис x_5. Минимальное симплексное отношение

$$\theta = \min_{a_{i5}>0}\{b_i/a_{i5}\} = 16/8 = 2$$

соответствует второй строке. Разрешающий элемент $a_{25} = 8$.

Переходим к следующему опорному плану $\mathbf{x}_2$. Для этого разрешающую строку $i = 2$ делим на разрешающий элемент $a_{25} = 8$. Разрешающий столбец $j_0 = 5$ заполняем нулями, кроме $a'_{25} = 1$. Остальные элементы симплексной таблицы (итерация 2) пересчитываем по правилу треугольника (прямоугольника) аналогично предыдущему.

Контроль вычислений производится так, как это делалось на предыдущей итерации.

Поскольку $\Delta_j \geqslant 0$, опорный план $\mathbf{x}_2$ оптимален. Итак,

$$\mathbf{x}^* = (7;\ 0;\ 0;\ 42;\ 2), \quad z(\mathbf{x}^*) = 72.$$

Пример 1.6. Решить с использованием искусственного базиса ЗЛП

$$\min Z = 3x_1 + 2x_2 + 3x_3;$$

$$\left.\begin{aligned} 2x_1 + \ x_2 + \ x_3 &\leqslant 2, \\ 3x_1 + 8x_2 + 2x_3 &\geqslant 8, \end{aligned}\right\}$$

$$x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_3 \geqslant 1.$$

Решение. Сведем задачу к каноническому виду. Получим:

$$\min Z = 3x_1 + 2x_2 + 3x_3 + 0\cdot x_4 + 0\cdot x_5 + 0\cdot x_6;$$

$$\left.\begin{array}{l} 2x_1 + \ x_2 + \ x_3 + \underline{x_4} \qquad\qquad = 2, \\ 3x_1 + 8x_2 + 2x_3 \qquad - x_5 \qquad = 8, \\ \qquad\qquad\quad x_3 \qquad\qquad - x_6 = 1, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,6}).$$

Первое ограничение имеет предпочтительную переменную, а второе и третье — нет. Поэтому вводим в них искусственные переменные w_1 и w_2. Приходим к M-задаче:

$$\min \bar{z} = 3x_1 + 2x_2 + 3x_3 + 0\cdot x_4 + 0\cdot x_5 + 0\cdot x_6 + Mw_1 + Mw_2;$$

$$\left.\begin{array}{l} 2x_1 + \ x_2 + \ x_3 + x_4 \qquad\qquad\qquad\qquad = 2, \\ 3x_1 + 8x_2 + 2x_3 \qquad - x_5 \qquad + \underline{w_1} \qquad = 8, \\ \qquad\qquad\quad x_3 \qquad\qquad - x_6 \qquad + \underline{w_2} = 1, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,6}), \quad w_1 \geqslant 0, \quad w_2 \geqslant 0.$$

Занесем условие M-задачи в симплексную таблицу 1.8.

Сделаем необходимые пояснения. В случае M-задачи индексную строку удобно разбить на две. В первой записываются свободные члены выражений $\Delta_0 = \mathbf{c}_Б \mathbf{A}_0$ и $\Delta_j = \mathbf{c}_Б \mathbf{A}_j - c_j$, а во второй — коэффициенты, содержащие M. Например, для нулевой итерации:

$$\Delta_0 = 0 \cdot 2 + 8M + 1 \cdot M = 0 + 9M,$$
$$\Delta_1 = 0 \cdot 2 + 3M + 0 \cdot M - 3 = -3 + 3M,$$
$$\Delta_2 = 0 \cdot 1 + 8M + 0 \cdot M - 2 = -2 + 8M, \ldots$$

По мере вывода из базиса искусственных переменных соответствующие им столбцы можно опускать.

Таблица 1.8

Номер итерации	БП	$\mathbf{c}_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	w_1	w_2	θ
				3	2	3	0	0	0	M	M	
0	x_4	0	2	2	1	1	1	0	0	0	0	$2/1 = 2$
	w_1	M	8	3	$\boxed{8}$	2	0	-1	0	1	0	$8/8 = 1$
	w_2	M	1	0	0	1	0	0	-1	0	1	—
	$\overline{z}_j - c_j$		0	-3	-2	-3	0	0	0	0	0	—
			$9M$	$3M$	$8M$	$3M$	0	$-M$	$-M$	0	0	—
1	x_4	0	1	13/8	0	3/4	1	1/8	0	—	0	$\frac{1}{3/4} = \frac{4}{3}$
	x_2	2	1	3/8	1	1/4	0	$-1/8$	0	—	0	$\frac{1}{1/4} = 4$
	w_2	M	1	0	0	$\boxed{1}$	0	0	-1	—	1	$1/1 = 1$
	$\overline{z}_j - c_j$		2	$-9/4$	0	$-5/2$	0	$-1/4$	0	—	0	—
			M	0	0	M	0	0	$-M$	—	0	—
2	x_4	0	1/4	13/8	0	0	1	1/8	3/4			
	x_2	2	3/4	3/8	1	0	0	$-1/8$	1/4			
	x_3	3	1	0	0	1	0	0	-1			
	$\overline{z}_j - c_j$		9/2	$-9/4$	0	0	0	$-1/4$	$-5/2$			

Так как в симплексной таблице (итерация 2) все оценки неположительны: $\Delta_j \leqslant 0$ $(j = \overline{1,6})$, то план оптимален.

Итак, $\mathbf{x}_2 = \mathbf{x}^* = (0; 3/4; 1; 1/4; 0; 0)$, $z(\mathbf{x}^*) = 9/2$.

Альтернативный оптимум (признак бесконечности множества оптимальных планов). Из геометрической интерпретации ЗЛП (см. рис. 1.4, *б*) следует, что если разрешающая прямая (плоскость, гиперплоскость) проходит через сторону (ребро, грань) многоугольника (многогранника, многогранной области) планов, то ЗЛП имеет бесконечное множество оптимальных планов. В этом случае говорят об альтернативном оптимуме. Как же определить, решая ЗЛП симплексным методом, единственный ли она имеет оптимальный план или бесконечное множество? На этот вопрос отвечает нижеследующая теорема.

Теорема 1.9. *Если в индексной строке последней симплексной таблицы, содержащей оптимальный план, имеется хотя бы одна нулевая оценка, соответствующая свободной переменной, то ЗЛП имеет бесконечное множество оптимальных планов.*

Доказательство. Пусть в оптимальном плане $\Delta_{j_0} = 0$, где j_0 принадлежит множеству индексов свободных переменных, а минимальное симплексное отношение соответствует строке i_0. Тогда, введя x_{j_0} в базис, получим новое значение целевой функции

$$\Delta_0' = \Delta_0 - \frac{\beta_{i_0}^* \Delta_{j_0}}{\alpha_{i_0 j_0}} = \Delta_0 - \frac{\beta_{i_0}^* \cdot 0}{\alpha_{i_0 j_0}} = \Delta_0.$$

Значение целевой функции при переходе к новому опорному плану не изменилось.

Если нулевых небазисных оценок в последней симплексной таблице окажется несколько, то, введя каждую из соответствующих переменных в базис, найдем оптимальные планы $\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_k^*$, для которых значение целевой функции будет одно и то же, т. е.

$$z(\mathbf{x}_1^*) = z(\mathbf{x}_2^*) = \ldots = z(\mathbf{x}_k^*).$$

Согласно второй части основной теоремы линейного программирования, в этом случае оптимальным будет любой план, являющийся их выпуклой линейной комбинацией, т. е. общее решение примет вид

$$\mathbf{x}^* = \lambda_1 \mathbf{x}_1^* + \lambda_2 \mathbf{x}_2^* + \ldots + \lambda_k \mathbf{x}_k^*, \quad \lambda_1 + \lambda_2 + \ldots + \lambda_k = 1,$$

$$\lambda_i \geqslant 0 \quad (i = \overline{1, k}).$$

Эта формула определяет бесконечное множество оптимальных планов. $\triangle$

С л е д с т в и е. *Если в индексной строке симплексной таблицы, содержащей оптимальный план, все оценки свободных переменных положительны, то найденный оптимальный план единственный.*

Признак неограниченности целевой функции. Как следует из геометрической интерпретации ЗЛП, возможны случаи, когда целевая функция при решении на максимум не ограничена сверху, а при решении на минимум не ограничена снизу (см. рис. 1.4, *г*). Такие случаи легко выявляются при решении задачи симплексным методом на основе следующих теорем.

Теорема 1.10. *Если в индексной строке симплексной таблицы ЗЛП на максимум содержится отрицательная оценка $\Delta_{j_0} < 0$, а в соответствующем столбце переменной x_{j_0} нет ни одного положительного элемента, то целевая функция на множестве допустимых планов задачи не ограничена сверху.*

Д о к а з а т е л ь с т в о. Если бы все оценки индексной строки Δ_j были неотрицательны, то опорный план был бы оптимальным. Пусть $\Delta_{j_0} < 0$. Тогда можно попытаться, не изменяя нулевых значений всех свободных переменных $x_{m+1}, \ldots, x_n$ в равенстве (1.95), кроме x_{j_0}, увеличить значение целевой функции Z за счет увеличения переменной $x_{j_0} > 0$. Полагая в равенствах (1.96) все x_j, кроме x_{j_0}, равными нулю, получаем

$$x_i = \beta_i - \alpha_{ij_0} x_{j_0} \quad (i = \overline{1, m}).$$

Так как $x_i \geqslant 0$, то

$$\beta_i - \alpha_{ij_0} x_{j_0} \geqslant 0 \quad (i = \overline{1, m}).$$

Пусть все $\alpha_{ij_0} \leqslant 0$. Тогда при любом $x_{i_0} \geqslant 0$ имеем $x_i \geqslant 0$. Что касается целевой функции Z, то, взяв в качестве x_{j_0} достаточно большое положительное число, вследствие того, что $\Delta_{j_0} < 0$, можно сделать значение

$$Z = \Delta_0 - (\Delta_{m+1} x_{m+1} + \ldots + \Delta_{j_0} x_{j_0} + \ldots + \Delta_n x_n)$$

как угодно большим.

В самом деле, так как $\Delta_{j_0} < 0$, при $x_{j_0} \to \infty$ имеем $Z = \Delta_0 - \Delta_{j_0} x_{j_0} \to \infty$, т.е. целевая функция Z не ограничена сверху. $\triangle$

Теорема 1.11. *Если в индексной строке симплексной таблицы ЗЛП на минимум содержится положительная оценка $\Delta_{j_0} > 0$, а в столбце переменной x_{j_0} нет ни одного положительного элемента, то на множестве допустимых планов целевая функция не ограничена снизу.*

Доказательство аналогично доказательству теоремы 1.10.

С экономической точки зрения неограниченность целевой функции ЗЛП свидетельствует только об одном: разработанная модель недостаточно точна. Бессмысленно говорить, например, о бесконечной прибыли. Типичными ошибками, приводящими к построению моделей такого рода, являются: 1) неполный учет ограничений, которые являются существенными в данной задаче; 2) небрежные оценки параметров, фигурирующих в ограничениях.

В заключение заметим, что при решении ЗЛП симплексным методом мы пользовались табличной записью условий задачи в форме табл. 1.5. На практике используются и другие формы таблиц. Очень компактной и наглядной является запись задачи (1.95) — (1.97) в форме табл. 1.9, называемой *жордановой*.

Таблица 1.9

БП	1	СП				
		$-x_{m+1}$	$\ldots$	$-x_{m+s}$	$\ldots$	$-x_n$
$x_1 =$	β_1	$\alpha_{1,m+1}$	$\ldots$	$\alpha_{1,m+s}$	$\ldots$	α_{1n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$x_k =$	β_k	$\alpha_{k,m+1}$	$\ldots$	$\alpha_{k,m+s}$	$\ldots$	α_{kn}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$x_m =$	β_m	$\alpha_{m,m+1}$	$\ldots$	$\alpha_{m,m+s}$	$\ldots$	α_{mn}
Z	Δ_0	Δ_{m+1}	$\ldots$	Δ_{m+s}	$\ldots$	Δ_n

При такой записи условий ЗЛП правила пересчета элементов таблиц при симплексном преобразовании будут следующими:

1) разрешающий элемент заменяется обратной величиной;

2) остальные элементы разрешающей строки делятся на разрешающий элемент;

3) остальные элементы разрешающего столбца делятся на разрешающий элемент и меняют свои знаки;

4) все прочие элементы вычисляются по правилу прямоугольника.

Сформулированные правила соответствуют модифицированным жордановым исключениям.

Пример 1.7. Решить ЗЛП

$$\max Z = 2x_1 + 3x_2;$$

$$\left.\begin{array}{l} x_1 + x_2 + x_3 = 5, \\ x_1 + 3x_2 + x_4 = 9, \\ x_1 + x_5 = 4, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,5}).$$

Решение. Записываем условия задачи в форме табл. 1.9 (см. табл. 1.10) и после двух итераций (табл. 1.10 — 1.12) получаем оптимальный план $\mathbf{x}^* = (3; 2; 0; 0; 1)$. При этом $\max Z = 12$.

Таблица 1.10

БП	1	СП	
		$-x_1$	$-x_2$
$x_3 =$	5	1	1
$x_4 =$	9	1	$\boxed{3}$
$x_5 =$	4	1	0
$Z =$	0	− 2	− 3

Таблица 1.11

БП	1	СП	
		$-x_1$	$-x_4$
$x_3 =$	2	$\boxed{2/3}$	− 1/3
$x_2 =$	3	1/3	1/3
$x_5 =$	4	1	0
$Z =$	9	− 1	1

Таблица 1.12

БП	1	СП	
		$-x_3$	$-x_4$
$x_1 =$	3		
$x_2 =$	2		
$x_5 =$	1		
$Z =$	12	3/2	1/2

Понятие о вырождении. Монотонность и конечность симплексного метода. Зацикливание. Очевидно, что базисный план ЗЛП, записанной в предпочтительном виде, невырожден, когда свободные члены всех уравнений положительны, и вырожден, если среди них имеются нули. Каноническую ЗЛП называют *невырожденной*, если все ее опорные планы невырожденные. Если среди опорных планов есть хотя бы один вырожденный, то задачу называют *вырожденной*.

Пусть ЗЛП решается на максимум. На двух последовательных итерациях значения целевой функции связаны соотношением (1.104):

$$\Delta_0^{(k+1)} = \Delta_0^{(k)} - \beta_{i_0}\Delta_{j_0}/\alpha_{i_0 j_0},$$

где $\beta_{i_0} \geqslant 0$; $\Delta_{j_0} \leqslant 0$; $\alpha_{i_0 j_0} > 0$.

Если задача невырожденная, то для любого шага $\beta_{i_0} > 0$, $\Delta_{j_0} \leqslant 0$. Отсюда $\Delta_0^{(k+1)} \geqslant \Delta_0^{(k)}$, т.е. значение целевой функции будет не хуже прежнего (монотонно возрастает). Аналогично, если задача решается на минимум, то, поскольку $\Delta_{j_0} \geqslant 0$, на любом шаге значения целевой функции на двух последовательных итерациях будут связаны неравенством $\Delta_0^{(k+1)} \leqslant \Delta_0^{(k)}$, т.е. целевая функция монотонно убывает. В этом и состоит монотонность симплексного метода.

Конечность алгоритма следует из конечного числа опорных планов ЗЛП (их не больше C_n^m).

Если ЗЛП вырожденная, то возможны случаи, когда $\beta_{i_0} = 0$. В этом случае значение целевой функции не улучшится. Предположим, что процесс продолжается без остановки, порождая бесконечную последовательность опорных планов. Вследствие конечности множества опорных планов в этой последовательности некоторые планы должны повторяться. В силу равенства для них целевой функции они должны лежать в одной гиперплоскости и быть вершинами выпуклого многогранника. Поэтому должна встретиться цепочка (цикл) $\mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_{k+l}$, в которой начальный и конечный планы совпадают, причем в силу монотонности метода

$$z(\mathbf{x}_k) = z(\mathbf{x}_{k+1}) = \ldots = z(\mathbf{x}_{k+l}).$$

Процесс продолжится неограниченно, если несколько последовательных вырождений приведут к образованию цикла. Такое явление называется *зацикливанием*. Зацикливание возможно только для вырожденных задач. Теория и практика показывают, что зацикливание возникает при весьма маловероятном сочетании условий. Известно лишь несколько специально разработанных (очень сложных) примеров, в процессе решения которых возникает зацикливание. Точный алгоритм вывода из цикла достаточно сложен. В простейшем случае при появлении цикла следует изменить последовательность вычислений путем изменения выбора разрешающего столбца.

Другое правило рекомендует изменить выбор разрешающей строки. Если в процессе симплексных преобразований появляется несколько минимальных симплексных отношений, то в качестве разрешающей выбирают ту строку, для которой будет наименьшим отношение элементов первого столбца к разрешающему. Если при этом снова оказывается несколько минимальных отношений, то составляются отношения элементов второго столбца к разрешающему, и так до тех пор, пока разрешающая строка не определится однозначно.

2. ДВОЙСТВЕННОСТЬ В ЛИНЕЙНОМ ПРОГРАММИРОВАНИИ

2.1. ПОНЯТИЕ ДВОЙСТВЕННОСТИ. ПОСТРОЕНИЕ ДВОЙСТВЕННЫХ ЗАДАЧ И ИХ СВОЙСТВА

Понятие двойственности для симметричных задач линейного программирования. Оно служит иллюстрацией одного из законов диалектики — закона единства и борьбы противоположностей. К аналогичным в математике относятся понятия "плюс" и "минус", "дифференцирование" и "интегрирование", в физике — "положительные" и "отрицательные" заряды, "северный" и "южный" магнитные полюса, в биологии — индивидуумы "мужского" и "женского" вида и т. д.

Понятие двойственности рассмотрим на примере задачи оптимального использования сырья. Пусть на предприятии решили рационально использовать отходы основного производства. В плановом периоде появились отходы сырья m видов в объемах b_i единиц ($i = \overline{1, m}$). Из этих отходов, учитывая специализацию предприятия, можно наладить выпуск n видов неосновной продукции. Обозначим через a_{ij} норму расхода сырья i-го вида на единицу j-й ($j = \overline{1, n}$) продукции, c_j — цена реализации единицы j-й продукции (реализация обеспечена). Неизвестные величины задачи: x_j — объемы выпуска j-й продукции, обеспечивающие предприятию максимум выручки.

Математическая модель задачи:

$$\max Z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n; \tag{2.1}$$

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leqslant b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leqslant b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leqslant b_m, \end{array}\right\} \tag{2.2}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}). \tag{2.3}$$

Предположим далее, что с самого начала при изучении вопроса об использовании отходов основного производства на предприятии появилась возможность реализации их некоторой организации. Необходимо установить прикидочные оценки (цены) на эти отходы. Обозначим их $y_1, y_2, \ldots, y_m$. Оценки должны быть установлены исходя из следующих требований, отражающих несовпадающие интересы предприятия и организации: 1) общую стоимость отходов сырья покупающая организация стремится минимизировать; 2) предприятие согласно уступить отходы только по таким ценам, при которых оно получит за них выручку, не меньшую той, что могло бы получить, организовав собственное производство. Эти требования формализуются в виде следующей ЗЛП.

Требование 1 покупающей организации — минимизация покупки:

$$\min f = b_1 y_1 + b_2 y_2 + \ldots + b_m y_m. \tag{2.4}$$

Требование 2 предприятия, реализующего отходы сырья, можно сформулировать в виде системы ограничений. Предприятие откажется от выпуска каждой единицы продукции первого вида, если

$$a_{11} y_1 + a_{21} y_2 + \ldots + a_{m1} y_m \geqslant c_1,$$

где левая часть означает выручку за сырье, идущее на единицу продукции первого вида, правая — ее цену.

Аналогичные рассуждения легко провести в отношении выпуска продукции каждого вида. Поэтому требование предприятия, реализующего отходы сырья, можно формализовать в виде следующей системы ограничений:

$$\left.\begin{array}{l} a_{11} y_1 + a_{21} y_2 + \ldots + a_{m1} y_m \geqslant c_1, \\ a_{12} y_1 + a_{22} y_2 + \ldots + a_{m2} y_m \geqslant c_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n} y_1 + a_{2n} y_2 + \ldots + a_{mn} y_m \geqslant c_n. \end{array}\right\} \tag{2.5}$$

По смыслу задачи оценки должны быть неотрицательными:

$$y_1 \geqslant 0, \quad y_2 \geqslant 0, \ \ldots, \ y_m \geqslant 0. \tag{2.6}$$

Переменные $y_i \quad (i = \overline{1,m})$ называются *двойственными оценками* или *объективно обусловленными оценками*. В зарубежной литературе их еще называют *теневыми ценами*.

Задачи (2.1) — (2.3) и (2.4) — (2.6) называются *парой взаимно двойственных ЗЛП*. Так как задачи (2.1) — (2.3) и (2.4) — (2.6) записаны в симметричной форме, их принято называть *парой симметричных двойственных задач*.

Пара взаимно двойственных симметричных задач в виде конечных сумм имеет вид:

прямая задача	*двойственная задача*
$\max Z = \sum\limits_{j=1}^{n} c_j x_j; \qquad (2.7)$	$\min f = \sum\limits_{i=1}^{m} b_i y_i; \qquad (2.10)$
$\sum\limits_{j=1}^{n} a_{ij} x_j \leqslant b_i \ (i=\overline{1,m}), \quad (2.8)$	$\sum\limits_{i=1}^{m} a_{ij} y_i \geqslant c_j \ (j=\overline{1,n}), \quad (2.11)$
$x_j \geqslant 0 \ (j=\overline{1,n}), \qquad (2.9)$	$y_i \geqslant 0 \ (i=\overline{1,m}) \qquad (2.12)$

или в матричной форме:

$\max Z = CX^T;$	$\min f = B^T Y;$
$AX^T \leqslant B,$	$A^T Y^T \geqslant C,$
$X \geqslant 0;$	$Y \geqslant 0.$

Можно показать, что если в качестве прямой принять задачу (2.4) — (2.6) об определении оптимальных оценок сырья, то двойственной к ней будет задача (2.1) — (2.3) об определении оптимального плана выпуска продукции.

Из моделей (2.1) — (2.3) и (2.4) — (2.6) непосредственно видно, что, имея математическую модель одной из этих задач, можно легко построить модель двойственной к ней задачи. Сопоставляя модели (2.1) — (2.3) и (2.4) — (2.6) пары двойственных задач, можно установить следующие взаимосвязи.

1. Если прямая задача — на максимум, то двойственная к ней — на минимум, и наоборот.

2. Коэффициенты c_j целевой функции прямой задачи являются свободными членами ограничений двойственной задачи.

3. Свободные члены b_i ограничений прямой задачи являются коэффициентами целевой функции двойственной.

4. Матрицы ограничений прямой и двойственной задач являются транспонированными друг к другу.

5. Если прямая задача — на максимум, то ее система ограничений представляется в виде неравенств типа $\leqslant$. Двойственная задача решается на минимум, и ее система ограничений имеет вид неравенств типа $\geqslant$.

6. Число ограничений прямой задачи равно числу переменных двойственной, а число ограничений двойственной — числу переменных прямой.

7. Все переменные в обеих задачах неотрицательны.

Пример 2.1. Исходя из специализации и своих технологических возможностей, предприятие может выпускать четыре вида продукции. Сбыт любого количества обеспечен. Для изготовления этой продукции используются трудовые ресурсы, полуфабрикаты и станочное оборудование. Общий объем ресурсов (в расчете на трудовую неделю), расход каждого ресурса на единицу выпускаемой продукции и цена, полученная за единицу продукции, приведены в табл. 2.1. Требуется определить план выпуска, доставляющий предприятию максимум выручки. Выполнить послеоптимизационный анализ решения и параметров модели.

Таблица 2.1

Ресурсы		Выпускаемая продукция				Объем ресурсов
		$П_1$	$П_2$	$П_3$	$П_4$	
P_1	Трудовые ресурсы, чел.-ч	4	2	2	8	4800
P_2	Полуфабрикаты, кг	2	10	6	0	2400
P_3	Станочное оборудование, станко-ч	1	0	2	1	1500
Прибыль, ден. ед.		65	70	60	120	

Решение. Пусть x_1, x_2, x_3, x_4 — объемы продукции $П_1$, $П_2$, $П_3$, $П_4$, планируемой к выпуску; Z — сумма ожидаемой выручки.

Математическая модель прямой задачи:

$$\max Z = 65x_1 + 70x_2 + 60x_3 + 120x_4;$$

$$\left.\begin{array}{l} 4x_1 + 2x_2 + 2x_3 + 8x_4 \leqslant 4800, \\ 2x_1 + 10x_2 + 6x_3 \leqslant 2400, \\ x_1 + 2x_3 + x_4 \leqslant 1500, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,4}).$$

Математическая модель двойственной задачи:

$$\min f = 4800y_1 + 2400y_2 + 1500y_3;$$

$$\left.\begin{aligned} 4y_1 + 2y_2 + y_3 &\geqslant 65,\\ 2y_1 + 10y_2 &\geqslant 70,\\ 2y_1 + 6y_2 + 2y_3 &\geqslant 60,\\ 8y_1 + y_3 &\geqslant 120, \end{aligned}\right\}$$

$$y_i \geqslant 0 \quad (i = \overline{1,3}).$$

Несимметричные двойственные задачи. Возьмем ЗЛП на максимум в канонической форме:

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{2.13}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = \overline{1,m}), \tag{2.14}$$

$$x_j \geqslant 0 \ (j = \overline{1,n}). \tag{2.15}$$

Для того чтобы записать двойственную задачу, представим ограничения-равенства (2.14) в виде системы равносильных неравенств:

$$\begin{aligned} &\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1,m}),\\ &\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad (i = \overline{1,m}). \end{aligned} \tag{2.16}$$

Умножив второе из равенств (2.16) на -1, получим задачу в симметричной форме:

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{2.17}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1,m}), \tag{2.18}$$

$$\sum_{j=1}^{n} (-a_{ij}) x_j \leqslant -b_i \quad (i = \overline{1,m}), \tag{2.19}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}). \tag{2.20}$$

Введем двойственные переменные y_i' и y_i'' для каждой системы ограничений (2.18), (2.19).

Двойственная задача (2.17) — (2.20) примет вид:

$$\min f = \sum_{i=1}^{m} b_i (y_i' - y_i''); \tag{2.21}$$

$$\sum_{i=1}^{m} a_{ij} (y_i' - y_i'') \geqslant c_j \quad (j = \overline{1, n}), \tag{2.22}$$

$$y_i' \geqslant 0,\ y_i'' \geqslant 0 \quad (i = \overline{1, m}). \tag{2.23}$$

Введем новые переменные

$$y_i' - y_i'' \equiv y_i. \tag{2.24}$$

Если переменную y_i считать соответствующей i-му ограничению (2.14) прямой задачи, то из соотношений (2.21) — (2.24) получим двойственную задачу в виде

$$\min f = \sum_{i=1}^{m} b_i y_i; \tag{2.25}$$

$$\sum_{i=1}^{m} a_{ij} y_i \geqslant c_j \quad (j = \overline{1, n}). \tag{2.26}$$

Из условий (2.23), (2.24) следует, что переменные y_i могут принимать как положительные, так и отрицательные значения, а также быть равными нулю, т. е. на знак этих переменных никаких ограничений не следует налагать. Задачи (2.13) — (2.15) и (2.25) — (2.26) представляют собой пару несимметричных двойственных задач. Так же формулируется двойственная задача в случае, когда в ограничения исходной задачи входят как неравенства, так и равенства:

прямая задача	*двойственная задача*
$\max Z = \sum\limits_{j=1}^{n} c_j x_j;$	$\min f = \sum\limits_{i=1}^{m} b_i y_i;$
$\sum\limits_{j=1}^{n} a_{ij} x_j \leqslant b_i \ (i=\overline{1,m_1}),$	$\sum\limits_{i=1}^{m} a_{ij} y_i \geqslant c_j \ (j=\overline{1,n_1}),$
$\sum\limits_{j=1}^{n} a_{ij} x_j = b_i \ (i=\overline{m_1+1,m}),$	$\sum\limits_{i=1}^{m} a_{ij} y_i = c_j \ (j=\overline{n_1+1,n}),$
$x_j \geqslant 0 \ (j=\overline{1,n_1};\ n_1 \leqslant n),$	$y_i \geqslant 0 \ (i=\overline{1,m_1};\ m_1 \leqslant m),$
x_j — любого знака $(j=\overline{n_1+1,n}),$	y_i — любого знака $(i=\overline{m_1+1,m}).$

Таким образом, двойственная задача со смешанными ограничениями составляется с соблюдением следующих дополнительных правил.

1. Если на переменную x_j прямой задачи наложено условие неотрицательности, то j-е условие системы ограничений двойственной задачи записывается в виде неравенства, и наоборот.

2. Если на переменную x_j прямой задачи не наложено условие неотрицательности, то j-е ограничение двойственной задачи записывается в виде строгого равенства.

3. Если в прямой задаче имеются ограничения равенства, то на соответствующие переменные двойственной задачи не налагается условие неотрицательности.

Очевидно, что задача, двойственная двойственной, совпадает с исходной. Поэтому безразлично, какую задачу принять в качестве прямой, а какую — двойственной. Следует говорить о паре взаимно двойственных задач.

Рассмотрим пару двойственных ЗЛП (2.7) — (2.9) и (2.10) — (2.12).

Теорема 2.1. *Для любых допустимых планов* $\mathbf{x}=(x_1;\dots;x_n)$ *и* $\mathbf{y}=(y_1;\dots;y_m)$ *прямой и двойственной ЗЛП справедливо неравенство* $z(\mathbf{x}) \leqslant f(\mathbf{y})$, *т. е.*

$$\sum_{j=1}^{n} c_j x_j \leqslant \sum_{i=1}^{m} b_i y_i. \qquad (2.27)$$

Доказательство. Учитывая неравенства (2.8) и (2.11), получаем

$$\sum_{j=1}^{n} c_j x_j \leqslant \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i \right) x_j = \sum_{i=1}^{m} y_i \left(\sum_{j=1}^{n} a_{ij} x_j \right) \leqslant \sum_{i=1}^{m} b_i y_i,$$

т. е. имеем неравенство (2.27), которое называется *основным неравенством теории двойственности.* $\triangle$

Экономическое содержание неравенства (2.27) состоит в том, что для любого допустимого плана производства $\mathbf{x}$ и любого допустимого вектора оценок ресурсов $\mathbf{y}$ общая созданная стоимость не превосходит суммарной оценки ресурсов.

***Теорема 2.2* (критерий оптимальности Канторовича).** *Если для некоторых допустимых планов $\mathbf{x}^*$ и $\mathbf{y}^*$ пары двойственных задач выполняется равенство $z(\mathbf{x}^*) = f(\mathbf{y}^*)$, то $\mathbf{x}^*$ и $\mathbf{y}^*$ являются оптимальными планами соответствующих задач.*

Доказательство. Согласно основному неравенству двойственности, для любого допустимого плана $\mathbf{x}$ прямой задачи и допустимого плана $\mathbf{y}^*$ двойственной справедливо неравенство $z(\mathbf{x}) \leqslant f(\mathbf{y}^*)$. Но по условию $z(\mathbf{x}^*) = f(\mathbf{y}^*)$. Отсюда в силу транзитивности отношений $\leqslant$ и $=$ получим $z(\mathbf{x}) \leqslant z(\mathbf{x}^*)$. Так как $\mathbf{x}$ — произвольный план, то $z(\mathbf{x}^*) = \max Z$, т. е. $\mathbf{x}^*$ — оптимальный план прямой ЗЛП. $\triangle$

Аналогично доказывается, что план $\mathbf{y}^*$ является оптимальным для двойственной задачи.

Экономическое содержание теоремы 2.2 состоит в том, что план производства $\mathbf{x}$ и вектор оценок ресурсов $\mathbf{y}$ являются оптимальными, если цена всей произведенной продукции и суммарная оценка ресурсов совпадают.

***Теорема 2.3* (малая теорема двойственности).** *Для существования оптимального плана любой из пары двойственных задач необходимо и достаточно существование допустимого плана для каждой из них.*

Доказательство. *Необходимость.* Пусть задачи двойственной пары имеют оптимальные планы $\mathbf{x}^*$ и $\mathbf{y}^*$. Это значит, что $\max z(\mathbf{x}) = z(\mathbf{x}^*)$ и $\min f(\mathbf{y}) = f(\mathbf{y}^*)$, т. е. $\mathbf{x}^*$ и $\mathbf{y}^*$ принадлежат области их допустимых планов. Соответствующие системы ограничений пары двойственных задач совместны, они имеют хотя бы по одному допустимому плану ($\mathbf{x}^*$ и $\mathbf{y}^*$). Необходимость доказана.

Достаточность. Пусть каждая из пары двойственных задач имеет допустимый план. Докажем, что они имеют оптимальные планы. Пусть $\mathbf{y}^*$ — допустимый план задачи (2.10) — (2.12). Тогда для любого допустимого плана $\mathbf{x}$ задачи (2.7) — (2.9), согласно основному неравенству теории двойственности (2.27), получим

$$z(\mathbf{x}) \leqslant f(\mathbf{y}^*). \tag{2.28}$$

Решая задачу (2.7) — (2.9) симплексным методом, получаем последовательность опорных планов $\mathbf{x}_0, \mathbf{x}_1, \ldots$, для которых $z(\mathbf{x}_0), z(\mathbf{x}_1), \ldots$ В силу неравенства (2.28) эта последовательность ограничена сверху. В ней найдется наибольшее значение целевой функции. Следовательно, существует допустимый план $\mathbf{x}^*$, для которого $z(\mathbf{x}) \leqslant z(\mathbf{x}^*)$. Аналогично доказывается, что $f(\mathbf{y}) \geqslant f(\mathbf{y}^*)$. $\triangle$

2.2. ОСНОВНЫЕ ТЕОРЕМЫ ДВОЙСТВЕННОСТИ И ИХ ЭКОНОМИЧЕСКОЕ СОДЕРЖАНИЕ

Теорема 2.4. *Если одна из двойственных задач имеет оптимальное решение, то и другая имеет оптимальное решение, причем экстремальные значения целевых функций равны $z(\mathbf{x}^*) = f(\mathbf{y}^*)$. Если одна из двойственных задач неразрешима вследствие неограниченности целевой функции на множестве допустимых решений, то система ограничений другой задачи противоречива.*

Доказательство. Приведем конструктивное доказательство теоремы. Рассмотрим пару симметричных двойственных задач (2.7) — (2.9) и (2.10) — (2.12). Вводя дополнительные переменные $x_{n+1}, \ldots, x_{n+m}$ в прямую задачу и $y_{m+1}, \ldots, y_{m+n}$ в двойственную и учитывая, что $\max Z = -\min(-Z)$, приводим модели задач к каноническому виду.

Прямая задача:

$$
\left.\begin{aligned}
&\min(-Z) = -c_1x_1 - c_2x_2 - \ldots - c_jx_j - \ldots - \\
&\qquad -c_sx_s - \ldots - c_nx_n; \\
&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1s}x_s + \ldots + \\
&\qquad +a_{1n}x_n + x_{n+1} = b_1, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ij}x_j + \ldots + a_{is}x_s + \ldots + \\
&\qquad +a_{in}x_n + x_{n+i} = b_i, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rj}x_j + \ldots + a_{rs}x_s + \ldots + \\
&\qquad +a_{rn}x_n + x_{n+r} = b_r, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mj}x_j + \ldots + a_{ms}x_s + \ldots + \\
&\qquad +a_{mn}x_n + x_{n+m} = b_m, \\
&\qquad x_j \geqslant 0 \quad (j = \overline{1, n+m}).
\end{aligned}\right\} \qquad (2.29)
$$

Двойственная задача:

$$
\left.\begin{aligned}
&\min f = b_1y_1 + b_2y_2 + \ldots + b_iy_i + \ldots + \\
&\qquad +b_ry_r + \ldots + b_my_m; \\
&-a_{11}y_1 - a_{21}y_2 - \ldots - a_{i1}y_i - \ldots - a_{r1}y_r - \ldots - \\
&\qquad -a_{m1}y_m + y_{m+1} = -c_1, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&-a_{1j}y_1 - a_{2j}y_2 - \ldots - a_{ij}y_i - \ldots - a_{rj}y_r - \ldots - \\
&\qquad -a_{mj}y_m + y_{m+j} = -c_j, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&-a_{1s}y_1 - a_{2s}y_2 - \ldots - a_{is}y_i - \ldots - a_{rs}y_r - \ldots - \\
&\qquad -a_{ms}y_m + y_{m+s} = -c_s, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&-a_{1n}y_1 - a_{2n}y_2 - \ldots - a_{in}y_i - \ldots - a_{rn}y_r - \ldots - \\
&\qquad -a_{mn}y_m + y_{m+n} = -c_n, \\
&\qquad y_i \geqslant 0 \quad (i = \overline{1, m+n}).
\end{aligned}\right\} \qquad (2.30)
$$

Теперь между переменными двойственных задач можно установить соответствие, сопоставляя свободным переменным одной задачи базисные переменные другой, и наоборот:

$$\begin{array}{cccccccccccccc} x_1 & x_2 & \dots & x_j & \dots & x_n & x_{n+1} & x_{n+2} & \dots & x_{n+i} & \dots & x_{n+m} \\ \updownarrow & \updownarrow & & \updownarrow & & \updownarrow & \updownarrow & \updownarrow & & \updownarrow & & \updownarrow \\ y_{m+1} & y_{m+2} & \dots & y_{m+j} & \dots & y_{m+n} & y_1 & y_2 & \dots & y_i & \dots & y_m \end{array}$$

Занесем условия задач (2.29) и (2.30) в симплексные таблицы. Но теперь будем рассматривать и такие базисы, в которых условие неотрицательности свободных членов уравнений не выполняется. Как следует из моделей (2.29) и (2.30), при установленном соответствии переменных обнаруживается следующее свойство коэффициентов: если x_{n+i} — базисные, x_j — свободные переменные исходной задачи, а y_i и y_{m+j} — соответствующие им свободные и базисные переменные двойственной задачи, то коэффициент, с которым член входит в уравнение, содержащее y_{m+j}, только зна́ком отличается от коэффициента, с которым x_j входит в уравнение, содержащее x_{n+i}. Более того, если для каждой из задач двойственной пары составить первые симплексные таблицы, то элементы столбца свободных членов одной таблицы только знаками будут отличаться от соответствующих элементов индексной строки другой.

Пусть, например, в исходной задаче преобразуем базис: переменную x_s введем в базис вместо переменной x_{n+r}. Пусть $a_{rs} \neq 0$. Будем сопровождать каждую такую замену соответствующей заменой базиса в двойственной задаче. Так как $x_s \leftrightarrow y_{m+s}$, $x_{n+r} \leftrightarrow y_r$, то в двойственной задаче переменную y_{m+s} нужно вывести из базиса и ввести вместо нее свободную переменную y_r.

Элементы a_{ij} прямой задачи преобразуются по правилу

$$a'_{ij} = \frac{a_{ij}a_{rs} - a_{rj}a_{is}}{a_{rs}}.$$

Соответствующие им элементы $-a_{ij}$ двойственной задачи преобразуются по правилу

$$(-a_{ij})' = \frac{(-a_{ij})(-a_{rs}) - (-a_{rj})(-a_{is})}{(-a_{rs})} = -\frac{a_{ij}a_{rs} - a_{rj}a_{is}}{a_{rs}}.$$

Аналогично преобразуются элементы столбца свободных членов и индексной строки симплексных таблиц, т. е. связь между элементами следующей пары симплексных таблиц сохраняется.

Если исходная задача имеет оптимальный план, то через конечное число шагов получится окончательная симплексная таблица, в которой в последней строке среди коэффициентов не будет ни одного положительного, а в столбце свободных членов — ни одного отрицательного элемента, кроме, быть может, значения целевой функции. Ей отвечает определенная симплексная таблица двойственной задачи, в которой в силу указанного соответствия между строками (столбцами) исходной задачи и столбцами (строками) двойственной столбец свободных членов не будет содержать ни одного отрицательного элемента, кроме, быть может, значения целевой функции, а в последней (индексной) строке не будет ни одного положительного коэффициента, так как в индексной строке одной таблицы стоят те же числа, что и в столбце свободных членов другой, только с противоположными знаками. Это значит, что для двойственной задачи также получен оптимальный план, причем минимальное значение функции $-Z$ только зна́ком отличается от минимального значения функции f, т. е. $\max Z = \min f$.

Отметим, что опорному оптимальному плану исходной задачи соответствует опорный оптимальный план двойственной, который оказывается записанным с противоположными знаками в индексной строке той же окончательной симплексной таблицы исходной задачи. Первая часть теоремы доказана.

Докажем вторую часть теоремы. Пусть функция Z не ограничена сверху, т. е. $-Z$ не ограничена снизу. Тогда на некотором этапе решения исходной задачи симплексным методом получится таблица, в которой для некоторого элемента индексной строки среди расположенных над ним элементов его столбца не будет ни одного положительного. В соответствующей строке сопутствующей симплексной таблицы двойственной задачи свободный член будет отрицательным, а среди коэффициентов при переменных не будет ни одного отрицательного. Уравнение, соответствующее такой строке, не может обратиться в тождество при неотрицательных значениях переменной. Это означает противоречивость системы ограничений двойственной задачи в области неотрицательных решений. $\triangle$

Теорема остается справедливой и для несимметричной пары двойственных задач.

Экономическое содержание первой теоремы двойственности состоит в следующем: если задача определения оптимального плана, максимизирующего выпуск продукции, разрешима, то разрешима и задача определения оценок ресурсов. Причем цена продукции, полученной при реализации оптимального плана, совпадает с суммарной оценкой ресурсов. Совпадение значений целевых функций для соответствующих планов пáры двойственных задач достаточно для того, чтобы эти планы были оптимальными. Это значит, что *план производства и вектор оценок ресурсов являются оптимальными тогда и только тогда, когда цена произведенной продукции и суммарная оценка ресурсов совпадают. Оценки выступают как инструмент балансирования затрат и результатов.* Двойственные оценки гарантируют рентабельность оптимального плана, т. е. равенство общей оценки продукции и ресурсов, и обусловливают убыточность всякого другого плана, отличного от оптимального. Двойственные оценки позволяют сопоставить и сбалансировать затраты и результаты системы.

Пример 2.2. По условиям примера 2.1 найти: 1) ассортимент выпускаемой продукции, обеспечивающий предприятию максимум реализации (максимум выручки); 2) оценки ресурсов, используемых при производстве продукции.

Решение. 1. Симплексным методом решаем прямую задачу, модель которой составлена в примере 2.1 (табл. 2.2).

После второй итерации все оценки оказались неотрицательными, значит, найденный опорный план является оптимальным: $\mathbf{x}^* = (0; 0; 400; 500; 0; 0; 200)$, $Z^* = z(\mathbf{x}^*) = 84\,000$.

Основные переменные $x_1^* = 0$, $x_2^* = 0$, $x_3^* = 400$, $x_4^* = 500$ показывают, что продукцию Π_1 и Π_2 выпускать нецелесообразно, а продукции Π_3 следует произвести 400 ед., Π_4 — 500 ед.

Дополнительные переменные показывают, что ресурсы P_1 и P_2 используются полностью ($x_5^* = x_6^* = 0$), а вот равенство $x_7^* = 200$ свидетельствует о том, что 200 единиц ресурса P_3 остались неиспользованными.

2. Выпишем из табл. 2.2 компоненты оптимального плана двойственной задачи (см. пример 2.1) — двойственные оценки. В канонической форме прямой задачи переменные x_1, x_2, x_3, x_4 являются свободными, а дополнительные переменные x_5, x_6, x_7 — базисными. В канонической форме двойственной задачи свободными будут y_1, y_2, y_3, а базисными — y_4, y_5, y_6, y_7. Соответствие между переменными примет вид

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_6 & y_7 & y_1 & y_2 & y_3 \end{array}$$

Учитывая это соответствие, выписываем из индексной строки последней (2-й) итерации компоненты искомого оптимального плана $\mathbf{y}^* = (15; 5; 0; 5; 10; 0; 0)$ — двойственные оценки.

Таблица 2.2

Номер итерации	БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
				65	70	60	120	0	0	0
0	x_5	0	4800	4	2	2	$\boxed{8}$	1	0	0
	x_6	0	2400	2	10	6	0	0	1	0
	x_7	0	1500	1	0	2	1	0	0	1
	$z_j - c_j$		0	−65	−70	−60	−120	0	0	0
1	x_4	120	600	1/2	1/4	1/4	1	1/8	0	0
	x_6	0	2400	2	0	$\boxed{6}$	0	0	1	0
	x_7	0	900	1/2	−1/4	7/4	0	−1/8	0	1
	$z_j - c_j$		72 000	−5	−40	−30	0	15	0	0
2	x_4	120	500	5/12	−1/6	0	1	1/8	−1/24	0
	x_3	60	400	1/3	5/3	1	0	0	1/6	0
	x_7	0	200	−1/12	−19/6	0	0	−1/8	−7/24	1
	$z_j - c_j$		84 000	5	10	0	0	15	5	0

В соответствии с теоремой 2.4 $\min f = \max Z = 84\,000$. Запишем это равенство в развернутой форме:

$$48\,000 \cdot 15 + 2400 \cdot 5 + 1500 \cdot 0 = 65 \cdot 0 + 70 \cdot 0 + 60 \cdot 400 + 120 \cdot 500.$$

Учитывая, что компоненты $y_1^* = 15$, $y_2^* = 5$, $y_3^* = 0$ представляют собой оценки ресурсов P_1, P_2, P_3, заключаем: при оптимальном плане оценка ресурсов, затраченных на выпуск продукции, совпадает с оценкой произведенной продукции. В этом состоит экономическое содержание теоремы 2.4.

Таким образом, оптимальность плана означает точное воплощение в оценке произведенной по этому плану продукции оценки всех израсходованных ресурсов, т. е. полное отсутствие непроизводительных затрат.

Теорема 2.5 **(о дополняющей нежесткости).** *Для того чтобы планы* $\mathbf{x}^*$ *и* $\mathbf{y}^*$ *пары двойственных задач были* *оптимальными, необходимо и достаточно выполнение условий:*

$$x_j^*\Big(\sum_{i=1}^{m} a_{ij}y_i^* - c_j\Big) = 0 \quad (j = \overline{1,n}), \tag{2.31}$$

$$y_i^*\Big(\sum_{j=1}^{n} a_{ij}x_j^* - b_i\Big) = 0 \quad (i = \overline{1,m}). \tag{2.32}$$

Доказательство. *Необходимость.* Пусть $\mathbf{x}^*$ и $\mathbf{y}^*$ — оптимальные планы пары двойственных задач:

$$\max Z = \sum_{j=1}^{n} c_j x_j; \qquad \min f = \sum_{i=1}^{m} b_i y_i;$$

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = \overline{1,m}), \qquad (2.33) \qquad \sum_{i=1}^{m} a_{ij}y_i \geqslant c_j,$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}), \qquad y_i \geqslant 0 \quad (i = \overline{1,m}).$$

Согласно теореме 2.4, для этих планов значения целевых функций совпадают: $z(\mathbf{x}^*) = f(\mathbf{y}^*)$, т. е.

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*. \tag{2.34}$$

Подставив в выражение (2.34) b_i из равенства (2.33), получим

$$\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ij}x_j^*\Big)y_i^* = \sum_{j=1}^{n} x_j^* \sum_{i=1}^{m} a_{ij}y_i^*,$$

откуда

$$\sum_{j=1}^{n} x_j^*\Big(\sum_{i=1}^{m} a_{ij}y_i^* - c_j\Big) = 0. \tag{2.35}$$

Поскольку $x_j^* \geqslant 0$ и $\sum\limits_{i=1}^{m} a_{ij}y_i^* - c_j \geqslant 0 \quad (j = \overline{1,n})$, то из равенства (2.35) следуют условия (2.31). Условия (2.32) доказываются аналогично.

Достаточность. Пусть для некоторых допустимых планов $\mathbf{x}^*$ и $\mathbf{y}^*$ выполняются условия (2.31). Докажем их оптимальность. Просуммировав равенства (2.31) по всем j от 1 до n и выполнив преобразования, противоположные предыдущим, получим выражение (2.34). Согласно критерию Канторовича, планы $\mathbf{x}^*$ и $\mathbf{y}^*$ являются оптимальными. $\triangle$

Мы доказали теорему для пары двойственных задач, одна из которых записана в канонической форме. Теорема, однако, остается справедливой для любой пары двойственных задач.

Условия (2.31), (2.32) называются *условиями дополняющей нежесткости.* Из них следует: *если какое-либо ограничение одной из задач ее оптимальным планом обращается в строгое неравенство, то соответствующая компонента оптимального плана двойственной задачи должна равняться нулю; если же какая-либо компонента оптимального плана одной из задач положительна, то соответствующее ограничение в двойственной задаче ее оптимальным планом должно обращаться в строгое равенство.*

Таким образом, если $\sum\limits_{j=1}^{n} a_{ij}x_j^* < b_i$, то $y_i^* = 0$; если $y_i^* > 0$, то $\sum\limits_{j=1}^{n} a_{ij}x_j^* = b_i$.

Точно так же, если $\sum\limits_{i=1}^{m} a_{ij}y_i^* > c_j$, то $x_j^* = 0$; если $x_j^* > 0$, то $\sum\limits_{i=1}^{m} a_{ij}y_i^* = c_j$. Экономически это означает, что если по некоторому оптимальному плану $\mathbf{x}^*$ производства расход i-го ресурса строго меньше его запаса b_i, то в оптимальном плане соответствующая двойственная оценка единицы этого ресурса равна нулю. Если же в некотором оптимальном плане оценок его i-я компонента строго больше нуля, то в оптимальном плане производства расход соответствующего ресурса равен его запасу. Отсюда следует вывод: *двойственные оценки могут служить мерой дефицитности ресурсов; дефицитный ресурс (полностью используемый по оптимальному плану производства) имеет положительную оценку, а ресурс избыточный (используемый не полностью) имеет нулевую оценку.*

Пример 2.3. В условиях примера 2.1 установить степень дефицитности используемых ресурсов и обосновать рентабельность оптимального плана.

Решение. В примере 2.2 найден оптимальный план $\mathbf{x}^* = (0; 0; 400; 500; 0; 0; 200)$ выпуска продукции. При этом плане третье ограничение прямой задачи (см. пример 2.1) выполняется как строгое неравенство: $0 + 2 \cdot 400 + 500 = 1300 < 1500$. Это означает, что расход ресурса P_3 меньше его запаса, т. е. ресурс P_3 избыточный. Именно поэтому в оптимальном плане $\mathbf{y}^* = (15; 5; 0; 5; 10; 0; 0)$ двойственной задачи оценка y_3^* этого ресурса равна нулю.

А вот оценки y_1^* и y_2^* ресурсов P_1 и P_2 выражаются положительными числами 15 и 5, что свидетельствует о дефицитности этих ресурсов: они при оптимальном плане используются полностью. В самом деле, ограничения по этим ресурсам (см. пример 2.1) выполняются как строгие равенства: $4 \cdot 0 + 2 \cdot 0 + 2 \cdot 400 + 8 \cdot 500 = 4800$, $2 \cdot 0 + 10 \cdot 0 + 6 \cdot 400 = 2400$.

Поскольку 15>5, ресурс P_1 считается более дефицитным, чем ресурс P_2. Это мы подтвердим более убедительно в примере 2.4.

На основе теоремы 2.5 нетрудно объяснить, почему не вошла в оптимальный план продукция Π_1 и Π_2: первое и второе ограничения двойственной задачи (см. пример 2.1) выполняются как строгие неравенства: $4 \cdot 15 + 2 \cdot 5 + 0 > 65$, $2 \cdot 15 + 10 \cdot 5 > 70$. Это означает, что оценки ресурсов, расходуемых на изготовление единицы продукции Π_1 и Π_2, превышают оценки единицы этой продукции. Понятно, что такую продукцию выпускать предприятию невыгодно. Что же касается продукции Π_3 и Π_4 ($x_3^* > 0$, $x_4^* > 0$), то выпуск ее оправдан, поскольку оценка израсходованных ресурсов совпадает с оценкой произведенной продукции: $2 \cdot 15 + 6 \cdot 5 + 2 \cdot 0 = 60$, $8 \cdot 15 + 0 = 120$.

Таким образом, в оптимальный план войдет только та продукция, которая выгодна предприятию, и не войдет убыточная продукция. В этом проявляется рентабельность оптимального плана.

Рассмотрим задачу определения оптимального плана выпуска продукции (см. пример 2.1). Предположим, что запасы ресурсов могут изменяться. Возникает вопрос: как эти изменения сказываются на экстремальном значении выручки предприятия? На этот вопрос отвечает следующая теорема.

***Теорема 2.6* (об оценках).** *Двойственные оценки показывают приращение функции цели, вызванное малым изменением свободного члена соответствующего ограничения задачи математического программирования, точнее,*

$$\frac{\partial z(\mathbf{x}^*)}{\partial b_i} = y_i^* \quad (i = \overline{1, m}). \tag{2.36}$$

Доказательство. Пусть имеем задачу математического программирования

$$\max Z = z(\mathbf{x}); \tag{2.37}$$

$$\varphi_i(\mathbf{x}) = b_i \quad (i = \overline{1, m}). \tag{2.38}$$

Перейдем от задачи на условный экстремум к задаче на безусловный экстремум. Для этого построим вспомогательную функцию Лагранжа:

$$L(\mathbf{x}, \mathbf{y}) = z(\mathbf{x}) + \sum_{i=1}^{m} y_i(b_i - \varphi(\mathbf{x})), \tag{2.39}$$

где y_i — неопределенные множители Лагранжа. Найдя частные производные функции Лагранжа по x_j и y_i и приравняв их нулю, получим необходимые условия экстремума функции Лагранжа:

$$\frac{\partial L(\mathbf{x}, \mathbf{y})}{\partial x_j} = \frac{\partial z(\mathbf{x})}{\partial x_j} - \sum_{i=1}^{m} y_i \frac{\partial \varphi_i(\mathbf{x})}{\partial x_j} = 0 \quad (j = \overline{1, n}), \tag{2.40}$$

$$\frac{\partial L(\mathbf{x}, \mathbf{y})}{\partial y_i} = b_i - \varphi_i(\mathbf{x}) = 0 \quad (i = \overline{1, m}). \tag{2.41}$$

Очевидно, что всякое экстремальное значение задачи (2.37), (2.38) удовлетворяет условиям (2.40), (2.41). В самом деле, пусть $\mathbf{x}^*$ — допустимый план, доставляющий функции цели $z(\mathbf{x})$ экстремальное значение. Так как $\mathbf{x}^*$ — допустимый план, то $\varphi_i(\mathbf{x}^*) = b_i$ и условия (2.41) выполняются. Следовательно, из функции Лагранжа (2.39) следует

$$L(\mathbf{x}^*, \mathbf{y}^*) = z(\mathbf{x}^*),$$

откуда

$$\frac{\partial L(\mathbf{x}^*, \mathbf{y}^*)}{\partial x_j} = \frac{\partial z(\mathbf{x}^*)}{\partial x_j} = 0,$$

т. е. условия (2.40) выполняются.

Каждая допустимая точка $\mathbf{x}^*$, в которой $z(\mathbf{x}^*)$ достигает экстремального значения, должна быть решением системы (2.40), (2.41). Это необходимые условия для отыскания экстремума.

Будем считать, что свободные члены системы ограничений (2.38) могут в некоторых пределах изменяться. Тогда в случае ЗЛП многогранник планов будет изменяться. Его крайние точки становятся функциями правых частей b_i, следовательно, экстремальное значение функции цели будет изменяться. Обозначим координаты крайних точек через $\mathbf{x}_1(\mathbf{b}) =$

$= \mathbf{x}_1(b_1, b_2, \ldots, b_m)$, $\mathbf{x}_2(\mathbf{b}) = \mathbf{x}_2(b_1, b_2, \ldots, b_m), \ldots, \mathbf{x}_n(\mathbf{b}) = \mathbf{x}_n(b_1, b_2, \ldots, b_m)$.

Рассмотрим теперь функцию Лагранжа как функцию, зависящую от вектора свободных членов $\mathbf{b}$:

$$L(\mathbf{b}) = z(\mathbf{x}(\mathbf{b})) + \sum_{i=1}^{m} y_i(b_i - \varphi_i(\mathbf{x}(\mathbf{b}))). \tag{2.42}$$

Дифференцируя функцию Лагранжа (2.42) по b_i, получаем

$$\frac{\partial L(\mathbf{b})}{\partial b_i} = \left(\frac{\partial z}{\partial x_j} - \sum_{i=1}^{m} y_i \frac{\partial \varphi_i}{\partial x_j}\right)\frac{\partial x_j}{\partial b_i} + \\ + \sum_{i=1}^{m}(b_i - \varphi_i(\mathbf{x}(\mathbf{b})))\frac{\partial y_i}{\partial b_i} + y_i. \tag{2.43}$$

Учитывая условия (2.40), (2.41), из равенства (2.43) получаем

$$\frac{\partial L(\mathbf{b})}{\partial b_i} = y_i \quad (i = \overline{1, m}).$$

Но для оптимального плана $L(\mathbf{x}^*, \mathbf{y}^*) = z(\mathbf{x}^*)$. Следовательно,

$$\frac{\partial z(\mathbf{x}^*)}{\partial b_i} = y_i^* \quad (i = \overline{1, m}). \quad \triangle$$

Перейдем к выяснению экономического содержания третьей теоремы двойственности. Для этого в равенстве (2.36) дифференциалы заменим приращениями. Получим $\Delta z(\mathbf{x}^*) \approx \approx y_i^* \Delta b_i$. При $\Delta b_i = 1$ имеем $\Delta z(\mathbf{x}^*) \approx y_i^*$. Отсюда величина двойственной оценки численно равна изменению целевой функции при изменении соответствующего свободного члена ограничений на единицу. В прикладных задачах двойственные оценки y_i^* часто называются *скрытыми, теневыми ценами* или *маргинальными оценками ресурсов.*

Пример 2.4. В условиях примера 2.1 изучить возможность дальнейшего совершенствования оптимального ассортимента выпускаемой продукции.

Решение. В примере 2.3 установлено, что ресурсы P_1 и P_2 являются дефицитными. В связи с этим на основе теоремы 2.6 можно утверждать, что на каждую единицу ресурса P_i, введенную в производство, будет получена дополнительная выручка $\Delta_i Z$, численно равная y_i^*. В самом деле, при $\Delta b_1 = 1$ получаем $\Delta_1 Z = y_1^* \Delta b_1 = 15 \cdot 1 = 15$. По тем же причинам каждая дополнительная единица ресурса P_2 обеспечит прирост $\Delta_2 Z$ выручки, равный 5 ден. ед. Теперь становится понятно, почему ресурс P_1 считается более дефицитным по сравнению с ресурсом P_2: он может содействовать получению большей выручки.

Что же касается избыточного ресурса P_3, то увеличение его запаса не приведет к росту выручки, поскольку $\Delta_3 Z = y_3^* \Delta b_3 = 0 \cdot \Delta b_3 = 0$. Из этих рассуждений следует, что оценки ресурсов позволяют совершенствовать план выпуска продукции.

Выясним экономический смысл оценок $y_4^*, y_5^*, y_6^*, y_7^*$ продукции $П_1$, $П_2$, $П_3$, $П_4$.

По оптимальному плану $\mathbf{x}^* = (0; 0; 400; 500; 0; 0; 200)$ выпускать следует продукцию $П_3$ и $П_4$. Оценки y_6^* и y_7^* этих видов продукции равны нулю. Что это означает практически, станет ясно, если представить оценки в развернутой записи:

$$y_6^* = (2y_1^* + 6y_2^* + 2y_3^*) - 60 = (2 \cdot 15 + 6 \cdot 5 + 2 \cdot 0) - 60 = 0,$$

$$y_7^* = (8y_1^* + y_3^*) - 120 = (8 \cdot 15 + 0) - 120 = 0.$$

Таким образом, нулевая оценка показывает, что эта продукция является неубыточной, поскольку оценка ресурсов, расходуемых на выпуск единицы такой продукции, совпадает с оценкой единицы изготовленной продукции.

Что же касается продукции $П_1$ и $П_2$, являющейся, как установлено в примере 2.3, убыточной, а потому и не вошедшей в оптимальный план, то для ее оценок $y_4^* = 5$ и $y_5^* = 10$ получаем:

$$y_4^* = (4y_1^* + 2y_2^* + y_3^*) - 65 = 70 - 65 = 5,$$

$$y_5^* = (2y_1^* + 10y_2^*) - 70 = 80 - 70 = 10.$$

Отсюда видно, что оценка убыточной продукции показывает, насколько будет снижать каждая изготовленная единица такой продукции достигнутый оптимальный уровень выручки.

В заключение необходимо сделать важное замечание. В примерах 2.1 — 2.4 мы выяснили экономическое содержание двойственных оценок применительно к условиям примера 2.1. Что касается других типов задач линейного программирования, то интерпретация их двойственных оценок может отличаться от приведенной выше. Иногда она настолько неочевидна, что представляет серьезную проблему, особенно в задачах, в которых ограничения имеют различные знаки ($\leqslant$, $=$, $\geqslant$) при неотрицательности правых частей.

2.3. ПРИМЕНЕНИЕ ОЦЕНОК В ПОСЛЕОПТИМИЗАЦИОННОМ АНАЛИЗЕ

Выше были рассмотрены вычислительные процедуры симплексного метода без использования и с использованием искусственного базиса. Процедуры эти сопряжены с большим объемом расчетной работы. Тем не менее каждому, кто изучает математическое программирование, весьма полезно знать (и уметь), как выполняется решение задачи вручную. Разумеется, вычислительная машина полностью освобождает от необходимости осуществлять вручную трудоемкий процесс решения ЗЛП. Для этого имеется широкий набор разнообразных пакетов прикладных программ (ППП), позволяющих выполнять соответствующие вычисления на ЭВМ как в пакетном режиме, так и в режиме обращения к ЭВМ с терминальных устройств. Однако в любом случае нужно помнить: для экономиста важно не столько найти оптимальный план задачи, сколько провести исчерпывающий экономико-математический анализ модели. Приняв непосредственное участие в исследованиях, связанных с разработкой линейных оптимизационных моделей, и решении даже простейших практических задач, можно убедиться в том, что бóльшую часть времени занимают построение модели, сбор необходимых данных и подготовка входной информации для машинных программ. Получив оптимальное решение пары взаимно двойственных задач на ЭВМ, следует подвергнуть его беспристрастной экспертной оценке. Ошибочно думать, что интерпретация на ЭВМ информации, содержащейся даже в последней симплексной таблице, возможна без ясного представления о том, почему и как работает симплексный метод. Всегда есть опасность получения ошибочных решений, за которыми стоит "авторитет" ЭВМ.

Если в итоге получается только оптимальное решение, то по существу мы ничего не получаем. Для анализа нужна хотя бы последняя результирующая симплексная таблица. Она насыщена важными данными, лишь небольшая часть которых составляет набор оптимальных значений переменных. Из результирующей симплексной таблицы либо непосредственно, либо путем простых дополнительных вычислений можно получить информацию относительно: 1) оптимального решения, статуса ресурсов и их ценности; 2) чувствительности оптимального решения к изменению объемов ресурсов, вариациям коэффициентов целевой функции и интенсивности потребления (нормам расхода) ресурсов. Как было показано выше, сведения, относящиеся к первому пункту, содержатся в оптимальном плане прямой $\mathbf{x}^*$ и двойственной $\mathbf{y}^*$ задач. Статус ресурса определяется тем, является ли он дефицитным или недефицитным, т. е. полное или частичное его использование предусматривает оптимальное решение задачи. Соответствующую информацию можно получить по значениям дополнительных (остаточных) переменных, положительные значения которых указывают на неполное использование соответствующих ресурсов. Говоря о ресурсах, фигурирущих в ЗЛП, подразумевают, что установлены некоторые максимальные пределы их величин, поэтому в соответствующих исходных ограничениях должен использоваться знак $\leqslant$ при положительности правых частей. Следовательно, ограничения типа $\geqslant b_i$ ($b_i > 0$) не могут рассматриваться как ограничения на ресурсы. Ограничения такого типа отражают то обстоятельство, что решение должно удовлетворять определенным требованиям, например обеспечению минимального спроса или минимальных отклонений от установленных структурных характеристик производства (сбыта).

Итак, анализ модели — не менее важный этап, чем получение оптимального решения по модели, а в некоторых случаях анализ дает больше информации для принятия решения, чем само решение. Анализ позволяет ответить на вопросы, связанные с повышением рентабельности предприятия, распределением ограниченных ресурсов между звеньями производства, увеличением выпуска продукции путем рациональной "расшивки" узких мест производства, согласованием интересов отдельных ячеек народного хозяйства и т. д. Мы в этом параграфе займемся послеоптимизационным анализом полученного решения. Анализу структуры модели и ее параметров посвящен следующий параграф.

Применение двойственных оценок в экономическом анализе покажем на примере модели задачи о рациональном использовании ресурсов (см. пример 2.1).

Выясним состав двойственной оценки. Для этого рассмотрим, например, первый ресурс (его запас $b_1 = 4800$). Он дефицитен. Увеличение запаса этого ресурса на единицу приведет к дополнительному выпуску продукции, что увеличит выручку на $y_1^* = 15$ ден. ед. За счет чего? Возьмем соответствующий столбец $A_5 = (\frac{1}{8};\ 0;\ -\frac{1}{8})^T$ табл. 2.2. Элементы его характеризуют изменение объемов выпуска продукции и остатка ресурса при увеличении первого ресурса на единицу, т. е. если заменить b_1 на $b_1' = b_1 + 1 = 4800 + 1 = 4801$, то выпуск продукции Π_4 $x_4^* = 500$ заменится на $x_4'^* = 500 + 1/8 = 500,125$; выпуск продукции Π_3 $x_3^* = 400$ — на $x_3'^* = 400 + 0 = 400$. Резерв же третьего ресурса сократится до $x_7'^* = 200 - 1/8 = 199,875$. При этом выручка возрастет на $120 \cdot \frac{1}{8} + 60 \cdot 0 + 0(-\frac{1}{8}) = 15$ ден. ед., что соответствует двойственной оценке первого ресурса. Аналогично при увеличении второго ресурса на единицу выручка возрастет на $120(-\frac{1}{24}) + 60 \cdot \frac{1}{16} + 0(-\frac{7}{24}) = 5$ ден. ед., что соответствует двойственной оценке второго ресурса. Полученные равенства и показывают, какие составляющие образуют двойственные оценки.

Рассмотрим проблему взаимозаменяемости ресурсов. Дефицитным мы назвали ресурс, который при реализации оптимального плана полностью используется. Если в ЗЛП имеются два или больше дефицитных ресурсов, то представляет интерес, каким количеством одного дефицитного ресурса можно заменить другой дефицитный ресурс, чтобы целевая функция принимала то же самое максимальное значение. В примере 2.1 первый и второй ресурсы дефицитны. Возникает, например, вопрос: если первый ресурс уменьшить на единицу, то насколько нужно увеличить второй дефицитный ресурс, чтобы максимальное значение целевой функции не изменилось? Введем понятие коэффициента взаимозаменяемости. Пусть b_i и b_k — объемы дефицитных ресурсов i-го и k-го вида. Обозначим отношение $\Delta b_k / \Delta b_i$ через η_{ki}, т. е.

$$\eta_{ki} = \frac{\Delta b_k}{\Delta b_i}.$$

Это отношение называется *коэффициентом взаимозаменяемости*. При $\Delta b_i = 1$ имеем $\eta_{ki} = \Delta b_k$, т. е. коэффициент взаимозаменяемости показывает, насколько нужно увеличить ресурс k-го вида, чтобы компенсировать уменьшение ресурса i-го вида на единицу. Заменив в соотношении $\eta_{ki} = \Delta b_k / \Delta b_i$ приращения на дифференциалы:

$$\frac{\Delta z}{\Delta b_i} \approx \frac{\partial z(x^*)}{\partial b_i} = y_i^*, \qquad \frac{\Delta z}{\Delta b_k} \approx \frac{\partial z(x^*)}{\partial b_k} = y_k^*,$$

получим

$$\eta_{ki} = \frac{\Delta z / y_k^*}{\Delta z / y_i^*} = \frac{y_i^*}{y_k^*}.$$

Если ресурс i-го вида уменьшить на Δb_i, то, увеличив ресурс k-го вида на $\Delta b_k = (y_i^* / y_k^*) \Delta b_i$, можно компенсировать уменьшение целевой функции. При этом, разумеется, сам оптимальный план изменится. В примере 2.1 дефицитны трудовые ресурсы и полуфабрикаты. Если бы трудовые ресурсы уменьшили на единицу, то связанное с этим падение выручки (на 15 ден. ед.) можно было бы компенсировать увеличением полуфабрикатов на

$$\Delta b_2 = \frac{y_1^*}{y_2^*} \Delta b_1 = \frac{15}{5} \cdot 1 = 3.$$

Следовательно, обеспечив полуфабрикаты в объеме $b_2' = b_2 + \Delta b_2 = 2400 + 3 = 2403$ (кг), можно получить с трудовыми ресурсами $b_1' = b_1 - \Delta b_1 = 4800 - 1 = 4799$ (чел.-ч) ту же выручку, что и при начальных ресурсах. В табл. 2.3 представлены значения коэффициентов взаимозаменяемости для примера 2.1. Знак ∞ означает, что заменить уменьшение на единицу одного ресурса никаким увеличением другого невозможно.

Таблица 2.3

k \ i	1	2	3
1	1	1/3	0
2	3	1	0
3	∞	∞	1

Проанализируем целесообразность расширения ассортимента выпускаемой продукции и установление цены на новую продукцию. Пусть в условиях примера 2.1 изучается вопрос о целесообразности выпуска продукции Π_5 с характеристиками, представленными в табл. 2.4.

Чтобы выпуск продукции Π_5 был оправдан, оценка ресурсов, затрачиваемых на изготовление единицы продукции Π_5, должна быть не менее цены $c_5 = 95$.

Таблица 2.4

Ресурсы	$П_5$
Трудовые P_1, чел.-ч	3
Полуфабрикаты P_2, кг	6
Станочное оборудование P_3, станко-ч	8
Цена единицы продукции, ден. ед.	95

Находим оценку затраченных ресурсов: $3 \cdot 15 + 6 \cdot 5 + 8 \cdot 0 = 75$. Поскольку $75 < 95$, выпускать продукцию $П_5$ целесообразно: каждая единица этой продукции принесет предприятию прибыль, равную $95 - 75 = 20$ ден. ед.

Аналогичным образом решается вопрос и об установлении нижней границы цены на новую продукцию. Если $a_{i,n+1}$ $(i = \overline{1,m})$ — нормы расхода соответствующих ресурсов P_i на единицу новой продукции, а ее цена равна c_{n+1}, то выпускать продукцию целесообразно при условии выполнения неравенства

$$a_{1,n+1}y_1^* + a_{2,n+1}y_2^* + \ldots + a_{m,n+1}y_m^* \leqslant c_{n+1}.$$

2.4. АНАЛИЗ ЛИНЕЙНЫХ МОДЕЛЕЙ

К анализу внутренней структуры модели прежде всего следует отнести вопросы анализа ее параметров на чувствительность. Под этим понимается влияние малого изменения параметров на значение целевой функции и на оптимальный план. Анализ на чувствительность дает возможность определить поведение модели в окрестности экстремума, а также пределы изменения исходных параметров, не изменяющих положение экстремума. Это позволяет сформулировать требования к точности исходных данных, а также упростить и уточнить модель, отбросив параметры, незначительно влияющие на конечный результат, и уточнив параметры, находящиеся в области высокой чувствительности. Анализ на чувствительность выявляет не только влияние на целевую функцию каждого из параметров задачи, но и одновременное изменение нескольких параметров. Ниже рассмотрен анализ коэффициентов целевой функции, ограничений по ресурсам и коэффициентов технологической матрицы.

Проанализируем коэффициенты целевой функции. Цель анализа коэффициентов целевой функции состоит в том, чтобы выяснить границы их изменения, в пределах которых оптимальный план исходной задачи не меняется. Пусть дана задача о рациональном использовании ресурсов (пример 2.1). Найдем нижнюю $\underline{c}_j$ и верхнюю $\overline{c}_j$ границы изменения коэффициента c_j, в пределах которых оптимальный план $\mathbf{x}^*$ задачи не меняется. Для этого в целевой функции дадим коэффициенту c_j приращение Δc_j и заменим c_j на $c_j + \Delta c_j$. Получим новую задачу, решение которой даст ответ на поставленный вопрос. Целесообразно рассмотреть три случая. Анализ коэффициентов целевой функции: 1) при свободных переменных; 2) при базисных переменных; 3) при одновременном изменении нескольких или всех коэффициентов целевой функции.

Случай 1. В примере 2.1 свободными переменными в оптимальном плане $\mathbf{x}^*$ являются x_1 и x_2. Поэтому исследуем пределы изменения c_1 и c_2. Начнем с коэффициента c_1. В целевой функции $\max Z = 65x_1 + 70x_2 + 60x_3 + 120x_4$ дадим коэффициенту c_1 приращение Δc_1. Получим задачу с целевой функцией:

$$\max Z' = (65 + \Delta c_1)x_1 + 70x_2 + 60x_3 + 120x_4, \quad \mathbf{x} \in \Omega.$$

Проделав симплексные преобразования с целевой функцией Z', придем к последней симплексной таблице 2.5, в которой в индексной строке оценка Δ_1^* будет заменена на $\Delta'^*_1 = \Delta_1^* - \Delta c_1$.

Таблица 2.5

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			$65 + \Delta c_1$	70	60	120	0	0	0
x_4	120	500	5/12	-1/6	0	1	1/8	-1/24	0
x_3	60	400	1/3	5/3	1	0	0	1/6	0
x_7	0	200	-1/12	-19/6	0	0	-1/8	-7/24	1
$z_j - c_j$		84 000	$5 + \Delta c_1$	10	0	0	15	5	0

Сравнение табл. 2.2 и 2.5 показывает, что при переходе от c_1 к $c_1 + \Delta c_1$ изменяется только оценка индексной строки Δ_1^*. Так как мы ищем пределы изменения для c_1, при которых оптимальность плана не нарушится, то $5 - \Delta c_1 \geqslant 0$. Отсюда

$-\infty < \Delta c_1 \leqslant 5$, т. е. $\underline{c}_1 = -\infty$, $\bar{c}_1 = c_1 + \Delta c_1 = 65 + 5 = 70$. Учитывая экономический смысл коэффициентов целевой функции (цена, прибыль и т. д.), в ряде задач следует считать $\underline{c} \geqslant 0$. Для нашей задачи $0 \leqslant c_1 \leqslant 70$.

Для коэффициента c_2 по аналогии имеем $\Delta'^*_2 = 10 - \Delta c_2 \geqslant 0$, т. е. $-\infty < \Delta c_2 \leqslant 10$. Отсюда $0 \leqslant c_2 \leqslant 80$.

Не представляет трудностей случай 1 обобщить для произвольной ЗЛП . Пусть $x_{j_0} \in$ (СП). Заменим c_{j_0} на $c_{j_0} + \Delta c_{j_0}$. Это приведет только к изменению в последней симплексной таблице коэффициента индексной строки $\Delta^*_{j_0}$, который примет значение $\Delta'^*_{j_0} = \Delta^*_{j_0} - \Delta c_{j_0}$. Так как план оптимален, то $\Delta^*_{j_0} - \Delta c_{j_0} \geqslant 0$, т. е. $\Delta c_{j_0} \leqslant \Delta^*_{j_0}$. Отсюда имеем:

$$\underline{c}_{j_0} = -\infty, \quad \bar{c}_{j_0} = c_{j_0} + \Delta^*_{j_0},$$

$$-\infty < c'_{j_0} \leqslant c_{j_0} + \Delta^*_{j_0}.$$

Если коэффициенты целевой функции по экономическим соображениям должны быть неотрицательными, то

$$0 \leqslant c'_{j_0} \leqslant c_{j_0} + \Delta^*_{j_0}.$$

Окончательно случай 1 можно сформулировать в виде следующего правила. Пусть $\mathbf{x}^*$ — оптимальный план ЗЛП о рациональном использовании ресурсов. Соответствующие оптимальному плану оценки обозначим $\Delta^*_j \geqslant 0$. Тогда если $j \in$ (СП), то при изменении коэффициентов в целевой функции в пределах $(-\infty; c_j + \Delta^*_j]$ целевая функция достигает экстремального значения в одной и той же крайней точке и ее значение не меняется.

Случай 2. В примере 2.1 базисными в оптимальном плане $\mathbf{x}^*$ являются переменные x_3 и x_4. Исследуем, например, пределы изменения коэффициента c_3. Для этого в целевой функции исходной задачи заменим коэффициент c_3 на $c_3 + \Delta c_3$. Получим

$$\max Z' = 65x_1 + 70x_2 + (60 + \Delta c_3)x_3 + 120x_4, \quad \mathbf{x} \in \Omega.$$

Отыскав оптимальный план, получим последнюю симплексную таблицу 2.6.

Таблица 2.6

БП	$c_Б$	A_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	$60+\Delta c_3$	120	0	0	0
x_4	120	500	5/12	−1/6	0	1	1/8	−1/24	0
x_3	$60+\Delta c_3$	400	1/3	5/3	1	0	0	1/6	0
x_7	0	200	−1/12	−19/6	0	0	−1/8	−7/24	1
$z_j - c_j$		$84\,000+$ $+400\Delta c_3$	$5+$ $+\frac{1}{3}\Delta c_3$	$10+$ $+\frac{5}{3}\Delta c_3$	0	0	15	$5+$ $+\frac{1}{6}\Delta c_3$	0
			y_4	y_5	y_6	y_7	y_1	y_2	y_3

План новой задачи будет оптимальным при неотрицательности оценок ${\Delta'}_j^*$:

$$5 + \frac{1}{3}\Delta c_3 \geqslant 0, \quad 10 + \frac{5}{3}\Delta c_3 \geqslant 0, \quad 5 + \frac{1}{6}\Delta c_3 \geqslant 0.$$

Отсюда $-6 \leqslant \Delta c_3 < \infty$. Границы же изменения коэффициента c_3 следующие: нижняя $\underline{c}_3 = c_3 + \Delta\underline{c}_3 = 60 - 6 = 54$, верхняя $\bar{c}_3 = c_3 + \Delta\bar{c}_3 = 60 + \infty = \infty$. Следовательно, исходный план остается оптимальным при $54 \leqslant c_3' < \infty$. Значение же целевой функции будет меняться по формуле $z(\mathbf{x}^*) = 84\,000 + 400\Delta c_3$. Как видно из табл. 2.6, при этом будут также меняться значения двойственных оценок:

$${y_2'}^* = 5 + \frac{1}{6}\Delta c_3, \quad {y_4'}^* = 5 + \frac{1}{3}\Delta c_3, \quad {y_5'}^* = 10 + \frac{5}{3}\Delta c_3.$$

Аналогично установим пределы изменения c_4. Последняя симплексная таблица для новой задачи

$$\max Z' = 65x_1 + 70x_2 + 60x_3 + (120 + \Delta c_4)x_4 \quad (\mathbf{x} \in \Omega)$$

имеет вид табл. 2.7.

Пределы изменения Δc_4 находим из условий ${\Delta'}_j^* \geqslant 0$, т. е.:

$$5+\frac{5}{12}\Delta c_4 \geqslant 0; \; 10-\frac{1}{6}\Delta c_4 \geqslant 0; \; 15+\frac{1}{8}\Delta c_4 \geqslant 0; \; 5-\frac{1}{24}\Delta c_4 \geqslant 0,$$

откуда $-12 \leqslant \Delta c_4 \leqslant 60$. Нижняя граница $\underline{c}_4 = c_4 + \underline{\Delta c}_4 = 120 - 12 = 108$, верхняя граница $\bar{c}_4 = c_4 + \overline{\Delta c}_4 = 120 + 60 = 180$. Итак, $108 \leqslant c_4' \leqslant 180$.

Таблица 2.7

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	$120+\Delta c_4$	0	0	0
x_4	$120+\Delta c_4$	500	5/12	−1/6	0	1	1/8	−1/24	0
x_3	60	400	1/3	5/3	1	0	0	1/6	0
x_7	0	200	−1/12	−19/6	0	0	−1/8	−7/24	1
$z_j - c_j$		$84\,000+$ $+500\Delta c_4$	$5+$ $+\frac{5}{12}\Delta c_4$	$10-$ $-\frac{1}{6}\Delta c_4$	0	0	$15+$ $+\frac{1}{8}\Delta c_4$	$5-$ $-\frac{1}{24}\Delta c_4$	0

Не представляет трудностей обобщить полученные результаты.

Замена в модели рационального использования ресурсов c_i, $i \in$ (БП), на $c_i + \Delta c_i$ приводит к изменению оценок, которые получаются из последней симплексной таблицы исходной задачи по формулам:

$$\Delta'^*_j = \Delta^*_j + \alpha_{ij}\Delta c_i \quad (i \in (\text{БП});\ j = \overline{1,n}\,),$$

где α_{ij} — элементы строки в последней симплексной таблице, соответствующей базисной переменной i. Так как план оптимален, то

$$\Delta'^*_j = \Delta^*_j + \alpha_{ij}\Delta c_i \geqslant 0.$$

Отсюда имеем:

$$\min \Delta c_i = \begin{cases} -\infty, & \text{если } \alpha_{ij} \leqslant 0,\ j \in (\text{СП}), \\ \max\left\{-\dfrac{\Delta^*_j}{\alpha_{ij}}\right\}, & \text{если } \alpha_{ij} < 0,\ j \in (\text{СП}), \end{cases}$$

$$\max \Delta c_i = \begin{cases} \min\left\{-\dfrac{\Delta^*_j}{\alpha_{ij}}\right\}, & \text{если } \alpha_{ij} < 0,\ j \in (\text{СП}), \\ +\infty, & \text{если } \alpha_{ij} \geqslant 0,\ j \in (\text{СП}). \end{cases}$$

Поэтому $\min \Delta c_i \leqslant \Delta c_i \leqslant \max \Delta c_i$. При изменении c_i, $i \in$ (БП), в пределах $[c_i + \min \Delta c_i; c_i + \max \Delta c_i]$ сохраняются значения всех базисных переменных x_i, но изменяется значение целевой функции по формуле

$$z'(\mathbf{x}^*) = \Delta'^*_0 = \Delta^*_0 + x^*_i \Delta c_i, \quad i \in (\text{БП}).$$

С л у ч а й 3. Пусть одновременно изменяется несколько коэффициентов в целевой функции. Найдем, например, область одновременного изменения параметров c_3 и c_4, для которой оптимальный план задачи не меняется. Имеем задачу

$$\max Z' = 65x_1 + 70x_2 + (60 + \Delta c_3)x_3 + (120 + \Delta c_4)x_4, \quad \mathbf{x} \in \Omega.$$

Как и в предыдущих случаях, нет необходимости решать ее сначала. Вся нужная информация для получения оценок новой задачи содержится в последней симплексной таблице исходной задачи. Получение оценок новой задачи показано в табл. 2.8.

Таблица 2.8

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	$60+$ $+\Delta c_3$	$120+$ $+\Delta c_4$	0	0	0
x_4	$120+$ $+\Delta c_4$	500	5/12	−1/6	0	1	1/8	−1/24	0
x_3	$60+$ $+\Delta c_3$	400	1/3	5/3	1	0	0	1/6	0
x_7	0	200	−1/12	−19/6	0	0	−1/8	−7/24	1
$z_j - c_j$		$84\,000+$ $+500\Delta c_4+$ $+400\Delta c_3$	$5+$ $+\frac{5}{12}\Delta c_4+$ $+\frac{1}{3}\Delta c_3$	$10-$ $-\frac{1}{6}\Delta c_4+$ $+\frac{5}{3}\Delta c_3$	0	0	$15+$ $+\frac{1}{8}\Delta c_4$	$5-$ $-\frac{1}{24}\Delta c_4+$ $+\frac{1}{6}\Delta c_3$	0

Так как оптимальный план $\mathbf{x}^* = (0; 0; 500; 400)$ исходной задачи должен быть оптимальным планом и новой задачи, оценки (см. табл. 2.8) должны остаться неотрицательными: $\Delta_j'^* \geqslant 0$, т. е.

$$\left.\begin{array}{rcrcrcl} 5 & + & \frac{5}{12}\Delta c_4 & + & \frac{1}{3}\Delta c_3 & \geqslant & 0, \\ 10 & - & \frac{1}{6}\Delta c_4 & + & \frac{5}{3}\Delta c_3 & \geqslant & 0, \\ 15 & + & \frac{1}{8}\Delta c_4 & & & \geqslant & 0, \\ 5 & - & \frac{1}{24}\Delta c_4 & + & \frac{1}{6}\Delta c_3 & \geqslant & 0. \end{array}\right\}$$

Эта система неравенств определяет область изменения Δc_3 и Δc_4. Ее нетрудно представить графически. Так как $c_3' = 60 + \Delta c_3$, $c_4' = 120 + \Delta c_4$, то, имея область изменения Δc_3 и Δc_4, можно построить область взаимосвязанного изменения коэффициентов c_3 и c_4.

Аналогично исследуется случай одновременного изменения всех коэффициентов целевой функции. Нетрудно записать соответствующую систему неравенств для определения области устойчивости оптимального плана:

для оценок основных переменных

$$\Delta_j^* + \sum_{i=1}^{m} \alpha_{ij} \Delta c_i - \Delta c_j \geqslant 0 \quad (j = \overline{1, n});$$

для оценок дополнительных переменных

$$\Delta_j^* + \sum_{i=1}^{m} \alpha_{ij} \Delta c_i \geqslant 0 \quad (j = \overline{n+2, n+m}).$$

Проанализируем ограничения по ресурсам. Здесь, как и при анализе коэффициентов целевой функции, целесообразно рассмотреть три случая: 1) анализ ограничений по ресурсам, которые при реализации оптимального плана исчерпываются полностью; 2) анализ ограничений по ресурсам, которые при реализации оптимального плана оказываются избыточными; 3) общий случай, когда одновременно могут изменяться несколько или все правые части системы ограничений.

Случай 1. Ресурсы, которые расходуются полностью, называются *дефицитными*. Методика их анализа аналогична анализу коэффициентов целевой функции. Пусть i-й ресурс дефицитный. Придадим его запасу b_i приращение Δb_i, т. е. заменим в i-м ограничении b_i на $b_i' = b_i + \Delta b_i$. Перейдем к новой задаче, в которой i-е ограничение заменено на $a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \leqslant b_i + \Delta b_i$. Решая новую задачу, считаем, что Δb_i — малое. Анализ ограничений по ресурсам предусматривает выяснение пределов изменения Δb_i, при которых структура опорного плана исходной задачи неизменна, т. е. перечень базисных переменных остается прежним, однако их значения будут изменяться. Так как при симплексных преобразованиях коэффициенты технологической матрицы и оценки преобразуются независимо от элементов вектор-столбца свободных членов $\mathbf{A}_0$, то опорный план новой задачи будет оптимальным при тех же неотрицательных оценках и тех же элементах вектор-столбцов $\mathbf{A}_j$ $(j = \overline{1, n+m})$, что и в исходной задаче. Так как запас b_i заменен на $b_i + \Delta b_i$, то значения координат опорного плана будут изменяться, однако в новой задаче в базис должны входить те же переменные, что и в исходной.

Все сказанное позволяет вычислять элементы вектор-столбца $\mathbf{A}_0$, имея данные соответствующей итерации исходной задачи. Значения вектор-столбца $\mathbf{A}_0$ новой задачи на последней итерации можно взять из данных соответствующей итерации исходной задачи. Обобщение непосредственных вычислений для исходной задачи и такой же задачи с правой частью для i-го ограничения, равной $b_i + \Delta b_i$, позволяет сделать вывод: значения базисных переменных новой задачи равны соответствующим значениям базисных переменных исходной задачи, сложенным с произведением соответствующих элементов вектор-столбца $\mathbf{A}_{n+i}$ (дополнительных переменных для i-го ресурса или, что то же самое, вектор-столбца двойственной переменной i-го ресурса) на приращение Δb_i. Фрагменты последних итераций исходной и новой задач представлены в табл. 2.9, 2.10.

Таблица 2.9

БП	$c_Б$	$\mathbf{A}_0$	$\cdots$	x_{n+i}	$\cdots$
			$\cdots$	c_{n+i}	$\cdots$
x_{l_1}	c_{l_1}	$x^*_{l_1}$	$\cdots$	$\alpha_{l_1,n+i}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{l_i}	c_{l_i}	$x^*_{l_i}$	$\cdots$	$\alpha_{l_i,n+i}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{l_m}	c_{l_m}	$x^*_{l_m}$	$\cdots$	$\alpha_{l_m,n+i}$	$\cdots$
$z_j - c_j$		Δ^*_0	$\cdots$	Δ^*_{n+i}	$\cdots$

Таблица 2.10

БП	$c_Б$	$\mathbf{A}_0$	$\cdots$	x_{n+i}	$\cdots$
			$\cdots$	c_{n+i}	$\cdots$
x_{l_1}	c_{l_1}	$x^*_{l_1} + \alpha_{l_1,n+i}\Delta b_i$	$\cdots$	$\alpha_{l_1,n+i}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{l_i}	c_{l_i}	$x^*_{l_i} + \alpha_{l_i,n+i}\Delta b_i$	$\cdots$	$\alpha_{l_i,n+i}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{l_m}	c_{l_m}	$x^*_{l_m} + \alpha_{l_m,n+i}\Delta b_i$	$\cdots$	$\alpha_{l_m,n+i}$	$\cdots$
$z_j - c_j$		$\Delta^*_0 + \Delta^*_{n+i}\Delta b_i$	$\cdots$	Δ^*_{n+i}	$\cdots$

Так как полученный в табл. 2.10 план оптимален, то он должен быть опорным, т. е. все его компоненты должны быть неотрицательными. Поэтому для определения границ изменения получаем следующую систему неравенств:

$$x_{l_i}^* + \alpha_{l_i,n+i}\Delta b_i \geqslant 0 \quad (i = \overline{1,m}),$$

откуда имеем:

$$\min \Delta b_i = \begin{cases} -\infty, & \text{если } \alpha_{i,m+1} \leqslant 0,\ \ i \in (\text{БП}), \\ \max\left\{-\dfrac{x_i^*}{\alpha_{i,m+i}}\right\}, & \text{если } \alpha_{i,m+i} > 0,\ \ i \in (\text{БП}), \end{cases}$$

$$\max \Delta b_i = \begin{cases} \min\left\{-\dfrac{x_i^*}{\alpha_{i,m+i}}\right\}, & \text{если } \alpha_{i,m+i} < 0,\ \ i \in (\text{БП}), \\ +\infty, & \text{если } \alpha_{i,m+i} \geqslant 0,\ \ i \in (\text{БП}). \end{cases}$$

Следовательно,

$$\min \Delta b_i \leqslant \Delta b_i \leqslant \max \Delta b_i,$$

$$b_i + \min \Delta b_i \leqslant b_i' \leqslant b_i + \max \Delta b_i.$$

В этих границах структура опорного плана не меняется, т. е. не меняется состав (наименование) выпускаемой продукции. Объемы же выпуска будут меняться, и их можно вычислить по формуле

$$x_i'^* = x_i^* + \alpha_{l_i,n+i}\Delta b_i.$$

Изменится также значение целевой функции:

$$\Delta_0'^* = \Delta_0^* + \Delta_{n+i}^*\Delta b_i.$$

Пределы изменения ограничений по ресурсам одновременно являются и пределами устойчивости двойственных оценок. При изменении дефицитного ресурса в пределах сохранения структуры оптимального плана значения основных и дополнительных переменных двойственной задачи не меняются.

Для рассматриваемого нами примера 2.1 дефицитными являются первый и второй ресурсы (труд и полуфабрикаты). Найдем пределы изменения, например, первого ресурса, при которых не изменяется структура плана выпускаемой продукции. Заменим запас $b_1 = 4800$ на $b_1' = b_1 + \Delta b_1 = 4800 + \Delta b_1$. Решение новой задачи получаем из последней таблицы решения исходной задачи (см. табл. 2.2) по сформулированному выше правилу. Это решение представлено в табл. 2.11.

Таблица 2.11

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	120	0	0	0
x_4	120	$500+\frac{1}{8}\Delta b_1$	5/12	−1/6	0	1	1/8	−1/24	0
x_3	60	$400+0\cdot\Delta b_1$	1/3	5/3	1	0	0	1/6	0
x_7	0	$200-\frac{1}{8}\Delta b_1$	−1/12	−19/6	0	0	−1/8	−7/24	1
z_j-c_j		$84\,000+15\Delta b_1$	5	10	0	0	15	5	0
							y_1	y_2	y_3

Для определения границ изменения Δb_1 имеем следующую систему неравенств:

$$\left.\begin{aligned} 500+\tfrac{1}{8}\Delta b_1 \geqslant 0,\\ 200-\tfrac{1}{8}\Delta b_1 \geqslant 0. \end{aligned}\right\}$$

Отсюда $-4000 \leqslant \Delta b_1 \leqslant 1600$, $\underline{b}_1 = 4800-4000 = 800$, $\bar{b}_1 = 4800+1600 = 6400$, т. е. $800 \leqslant b_1' \leqslant 6400$.

Аналогично находим границы изменения второго ресурса. Для Δb_2 имеем следующую систему неравенств:

$$\left.\begin{aligned} 500 - \tfrac{1}{24}\Delta b_2 \geqslant 0,\\ 400 + \tfrac{1}{6}\Delta b_2 \geqslant 0,\\ 200 - \tfrac{7}{24}\Delta b_2 \geqslant 0, \end{aligned}\right\}$$

откуда $-2400 \leqslant \Delta b_2 \leqslant 4800/7$, $0 \leqslant b_2' \leqslant 216\,000/7$.

Случай 2. Исследование границ изменения ресурсов, которые расходуются не полностью, проводится по той же схеме, что и для случая 1. Так как для i-го ресурса дополнительная переменная x_{n+i} входит в базис и показывает величину ресурса, избыточного при реализации оптимального плана, то, уменьшив этот ресурс на Δb_2, мы не изменим структуру оптимального плана. Увеличение же его на сколь угодно большую величину также не изменит структуру плана, так как он и так избыточен. Следовательно:

$$x_{n+i}^* + \Delta b_i \geqslant 0, \quad n+i \in (\text{БП}), \quad y_i^* = 0,$$

откуда

$$-x_{n+i}^* \leqslant \Delta b_i < \infty, \quad b_i - x_{n+i}^* \leqslant b_i' < \infty.$$

Для примера 2.1 избыточным является третий ресурс (станочное оборудование). Дополнительная переменная $x_{4+3}^* = x_7^* = 200$ показывает, что в оптимальном плане 200 станко-ч не использовано. Следовательно, $-200 \leqslant \Delta b_3 < < \infty$, $1500 - \Delta \underline{b}_3 \leqslant b_3' \leqslant 1500 + \Delta \bar{b}_3$, т. е. $1300 \leqslant b_3' < \infty$.

Случай 3. Рассмотрим, например, первый и второй ресурсы. Дадим их объемам приращения Δb_1 и Δb_2. Последняя симплексная таблица примет вид табл. 2.12.

Таблица 2.12

БП	св	A_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	120	0	0	0
x_4	120	$500 + \frac{1}{8}\Delta b_1 - \frac{1}{24}\Delta b_2$	5/12	−1/6	0	1	1/8	−1/24	0
x_3	60	$400 + 0 \cdot \Delta b_1 + \frac{1}{6}\Delta b_2$	1/3	5/3	1	0	0	1/6	0
x_7	0	$200 - \frac{1}{8}\Delta b_1 - \frac{7}{24}\Delta b_2$	−1/12	−19/6	0	0	−1/8	−7/24	1
$z_j - c_j$		$84\,000 + 15\Delta b_1 + 5\Delta b_2$	5	10	0	0	15	5	0
							y_1	y_2	y_3

Чтобы план оставался оптимальным и опорным, необходимо соблюдать условия:

$$\left.\begin{aligned} 500 + \tfrac{1}{8}\Delta b_1 - \tfrac{1}{24}\Delta b_2 &\geqslant 0, \\ 400 + 0 \cdot \Delta b_1 + \tfrac{1}{6}\Delta b_2 &\geqslant 0, \\ 200 - \tfrac{1}{8}\Delta b_1 - \tfrac{7}{24}\Delta b_2 &\geqslant 0. \end{aligned}\right\}$$

Эту область изменения Δb_1 и Δb_2 можно изобразить графически. Воспользовавшись преобразованием $b_1' = 4800 + \Delta b_1$, $b_2' = 2400 + \Delta b_2$, можно построить соответствующую область взаимосвязанного изменения первого и второго ресурсов.

В общем случае область взаимосвязанного изменения Δb_i определится системой неравенств

$$x_i^* + \sum_{k=1}^{m} \alpha_{i,n+k} \Delta b_k \geqslant 0, \quad i \in (\text{БП}).$$

Проанализируем коэффициенты технологической матрицы. В силу громоздкости математических выкладок в общем виде методику анализа коэффициентов технологической матрицы покажем на примере 2.1. Возможны четыре случая, соответствующих переменным (свободным или базисным) и ограничениям по ресурсам (избыточным или дефицитным).

С л у ч а й 1 (анализ коэффициентов технологической матрицы при свободных переменных в ограничениях по ресурсам, которые являются избыточными). Этому случаю в примере 2.1 удовлетворяют коэффициенты $a_{31} = 1$ и $a_{32} = 0$ (свободными являются переменные x_1 и x_2, а избыточен третий ресурс). Так как продукция $П_1$ и $П_2$ убыточная и третий ресурс не сдерживает выпуск прибыльных видов продукции $П_3$ и $П_4$, вошедших в оптимальный ассортимент выпуска, то при произвольном изменении этих коэффициентов остальные характеристики оптимального плана не изменятся, т. е.

$$0 \leqslant a_{31} < \infty, \quad 0 \leqslant a_{32} < \infty.$$

К этому заключению можно прийти, применив общую методику, придав этим коэффициентам приращения и проделав соответствующие преобразования, ведущие к оптимальному плану. Вывод по случаю 1 таков: если коэффициенты технологической матрицы a_{ij} $(j \in (\text{СП});\ \ y_i^* = 0)$ соответствуют свободным переменным и находятся в ограничениях, ресурс которых избыточен, то к их точности никаких требований можно не предъявлять.

С л у ч а й 2 (анализ коэффициентов технологической матрицы при базисных переменных в ограничениях по ресурсам, которые избыточны). В примере 2.1 этому условию удовлетворяют коэффициенты $a_{33} = 2$, $a_{34} = 1$. Уменьшение норм расхода вызовет увеличение резерва по третьему ресурсу, поэтому на структуру оптимального плана не повлияет. Увеличение же норм расхода a_{33} и a_{34} повлечет уменьшение резерва по третьему ресурсу. При значительном увеличении норм расхода a_{33} и a_{34} третий ресурс может оказаться дефицитным и нарушится структура оптимального плана. Для того чтобы найти пределы изменения норм, при которых оптимальный план задачи остается прежним, заменим a_{ij} на $a_{ij} + \Delta a_{ij}$. Получим новую задачу. Ее решение можно выписать из последней симплексной таблицы исходной задачи. Чтобы сформулировать правило, позволяющее это сделать, решим новую задачу симплексным методом и сравним последнюю симплексную таблицу исходной задачи с последней таблицей решения новой задачи.

При замене коэффициента a_{ij} (соответствующего базисной переменной в ограничении с избыточным ресурсом) на $a_{ij} + \Delta a_{ij}$ в последней симплексной таблице новой задачи изменению подвергается только i-я строка: ее элементы равны соответствующим элементам последней симплексной таблицы решения исходной задачи за вычетом произведения элементов строки, соответствующей этой базисной переменной, на Δa_{ij}. Для коэффициента a_{33} получим табл. 2.13.

Таблица 2.13

БП	$c_Б$	A_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	120	0	0	0
x_4	120	500	5/12	−1/6	0	1	1/8	−1/24	0
x_3	60	400	1/3	5/3	1	0	0	1/6	0
x_7	0	$200-$ $-400a_{33}$	$-1/12-$ $-\frac{1}{3}a_{33}$	$-19/6-$ $-\frac{5}{3}a_{33}$	0	0	$-1/8-$ $-0\cdot a_{33}$	$-7/24-$ $-\frac{1}{6}a_{33}$	1
$z_j - c_j$		84 000	5	10	0	0	15	5	0

Собственно, для выявления границ устойчивости Δa_{33} новые коэффициенты технологической матрицы a'_{33} не нужны. Как видно из табл. 2.13, решение сохранится оптимальным и опорным, если $200 - 400\Delta a_{33} \geqslant 0$. Отсюда $-\infty < \Delta a_{33} \leqslant 0{,}5$. Поскольку $a_{ij} \geqslant 0$, то $\underline{a}_{33} = 0$, $\overline{a}_{33} = a_{33} + \Delta\overline{a}_{33} = 2 + 0{,}5 = 2{,}5$.

Следовательно, при $0 \leqslant a'_{33} \leqslant 2{,}5$ оптимальный план не изменяется. Изменится только величина резерва по избыточному ресурсу по формуле $x_7 = x_{4+3} = 200 - 400\Delta a_{33}$.

В общем виде имеем условие

$$x'^{*}_{m+i} = x^{*}_{m+i} - x^{*}_{j}\Delta a_{ij} \geqslant 0.$$

Отсюда, учитывая, что $a_{ij} \geqslant 0$, имеем

$$0 \leqslant a'_{ij} \leqslant a_{ij} + \frac{x^{*}_{m+i}}{x^{*}_{j}}.$$

Для коэффициента $a_{34} = 1$ получаем $0 \leqslant a'_{34} \leqslant 1 + 200/500$, т. е. $0 \leqslant a'_{34} \leqslant 1{,}4$.

При одновременном изменении всех коэффициентов технологической матрицы при базисных переменных в ограничениях по ресурсам, имеющим резервы, оптимальный план не изменится при выполнении системы неравенств

$$x_{n+i}^* - \sum_{j\in(\text{БП})} x_j^* \Delta a_{ij} \geqslant 0, \quad n+i \in (\text{БП}).$$

Случай 3 (анализ коэффициентов при свободных переменных в ограничениях, ресурсы которых используются полностью). В нашем примере (см. табл. 2.2) такими коэффициентами являются $a_{11} = 4$, $a_{12} = 2$, $a_{21} = 2$ и $a_{22} = 10$. Рассмотрим, например, коэффициент $a_{21} = 2$, который заменим на $a_{21} + \Delta a_{21} = 2 + \Delta a_{21}$. Решение новой задачи представлено в табл. 2.14.

Таблица 2.14

БП	$c_Б$	A_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	120	0	0	0
x_4	120	500	$\frac{5}{12} - \frac{1}{24}\Delta a_{21}$	-1/6	0	1	1/8	-1/24	0
x_3	60	400	$\frac{1}{3} + \frac{1}{6}\Delta a_{21}$	5/3	1	0	0	1/6	0
x_7	0	200	$-\frac{1}{12} - \frac{7}{24}\Delta a_{21}$	-19/6	0	0	-1/8	-7/24	1
$z_j - c_j$		84 000	$5 + 5\Delta a_{21}$	10	0	0	15	5	0

Отсюда следует: чтобы получить последнюю симплексную таблицу новой задачи, нужно к элементам j-го столбца в последней симплексной таблице исходной задачи прибавить произведение элементов столбца, соответствующего двойственной переменной для i-го ресурса, на Δa_{ij}. Как видно из табл. 2.14, для анализа коэффициента a_{ij} $(y_i^* > 0;\ j \in (\text{СП}))$ нужна только оценка Δ_j^*. Решение остается оптимальным при $\Delta_j^* \geqslant 0$. Для нашего примера $\Delta_1 = 5 + 5\Delta a_{21} \geqslant 0$. Отсюда $\Delta a_{21} \geqslant -1$, т.е. $-1 \leqslant \Delta a_{21} < +\infty$. Следовательно, $\underline{a}_{21} = 2 - 1 = 1$, $\overline{a}_{21} = 1 + \infty = +\infty$. При изменении коэффициента a_{21} в пределах $[1; +\infty)$ оптимальный план задачи не изменится.

Для общего случая $(y_i^* > 0,\ j \in (\text{СП}))$ имеем:

$$\Delta \underline{a}_{ij} = -\frac{y_{m+j}^*}{y_i^*}, \quad \Delta \overline{a}_{ij} = +\infty,$$

$$a_{ij} - \frac{y^*_{m+j}}{y^*_i} \leqslant a^*_{ij} < +\infty.$$

При одновременном изменении нескольких или всех коэффициентов при свободных переменных в ограничениях, не имеющих резерва по ресурсам, оптимальный план остается неизменным при условии

$$y^*_{m+j} + \sum_{m+i \in (\text{СП})} y^*_i \Delta a_{ij} \geqslant 0.$$

Случай 4 (анализ коэффициентов при базисных переменных в ограничениях, ресурсы которых используются полностью). Эти коэффициенты несут наибольшую нагрузку в формировании структуры оптимального плана. Всякое их изменение влияет на координаты оптимального плана и значение целевой функции. Поэтому возникает задача установления пределов изменения, в которых сохраняется хотя бы структура оптимального плана (т. е. наименования продукции, объемы выпуска которой в оптимальном плане отличны от нуля). В рассматриваемом примере 2.1 такими коэффициентами являются $a_{13} = 2$, $a_{14} = 8$, $a_{23} = 6$, $a_{24} = 0$.

Рассмотрим, например, коэффициент $a_{14} = 8$. Все вычисления представлены в табл. 2.15. При рассуждениях и расчетах предполагается, что приращение Δa_{14} достаточно малое. Из табл. 2.15 следует, что структура исходного плана не изменится, если для любых ${\Delta'_j}^* \geqslant 0$ и ${x'_i}^* \geqslant 0$ $i \in$ (БП). Заметим, что $8 + \Delta a_{14} > 0$. В противном случае, например, ${\Delta'_5}^* = 120/(8+\Delta a_{14})$ было бы отрицательным, что противоречит оптимальности плана (Δa_{14} — достаточно малое). Отсюда следует система неравенств:

$$\left.\begin{aligned} 7200 + 1500\Delta a_{14} &\geqslant 0,\\ 1600 + 2300\Delta a_{14} &\geqslant 0,\\ 40 - 45\Delta a_{14} &\geqslant 0,\\ 80 + 30\Delta a_{14} &\geqslant 0,\\ 40 + 10\Delta a_{14} &\geqslant 0, \end{aligned}\right\}$$

откуда получаем $168/23 \leqslant a_{14} \leqslant 80/9$.

Таблица 2.15

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_6	x_7
			65	70	60	120	0	0	0
x_5	0	4800	4	2	2	$\boxed{8+\Delta a_{14}}$	1	0	0
x_6	0	2400	2	10	6	0	0	1	0
x_7	0	1500	1	0	2	1	0	0	1
$z_j - c_j$		0	-65	-70	-60	-120	0	0	0
x_4	120	$\frac{4800}{8+\Delta a_{14}}$	$\frac{4}{8+\Delta a_{14}}$	$\frac{2}{8+\Delta a_{14}}$	$\frac{2}{8+\Delta a_{14}}$	1	$\frac{1}{8+\Delta a_{14}}$	0	0
x_6	0	2400	2	10	$\boxed{6}$	0	0	1	0
x_7	0	$\frac{7200+1500\Delta a_{14}}{8+\Delta a_{14}}$	$\frac{4+\Delta a_{14}}{8+\Delta a_{14}}$	$-\frac{2}{8+\Delta a_{14}}$	$\frac{14-2\Delta a_{14}}{8+\Delta a_{14}}$	0	$\frac{1}{8+\Delta a_{14}}$	0	1
$z_j - c_j$		$\frac{576000}{8+\Delta a_{14}}$	$\frac{-40-65\Delta a_{14}}{8+\Delta a_{14}}$	$\frac{-320-70\Delta a_{14}}{8+\Delta a_{14}}$	$\frac{-240-60\Delta a_{14}}{8+\Delta a_{14}}$	0	$\frac{120}{8+\Delta a_{14}}$	0	1
x_4	120	$\frac{4000}{8+\Delta a_{14}}$	$\frac{10}{24+3\Delta a_{14}}$	$\frac{4}{24+3\Delta a_{14}}$	0	0	$\frac{1}{8+\Delta a_{14}}$	$\frac{-1}{24+3\Delta a_{14}}$	0
x_3	60	400	$\frac{1}{3}$	$\frac{5}{3}$	1	0	0	$\frac{1}{6}$	0
x_7	0	$\frac{1600+2300\Delta a_{14}}{8+\Delta a_{14}}$	$\frac{-2+5\Delta a_{14}}{24+3\Delta a_{14}}$	$\frac{-76+10\Delta a_{14}}{24+3\Delta a_{14}}$	0	0	$\frac{1}{8+\Delta a_{14}}$	$\frac{-7+\Delta a_{14}}{24+3\Delta a_{14}}$	1
$z_j - c_j$		$\frac{672+24000}{8+\Delta a_{14}}$	$\frac{40-45\Delta a_{14}}{8+\Delta a_{14}}$	$\frac{80+30\Delta a_{14}}{8+\Delta a_{14}}$	0	0	$\frac{120}{8+\Delta a_{14}}$	$\frac{40+10\Delta a_{14}}{8+\Delta a_{14}}$	0

Общие формулы для определения любого коэффициента новой последней симплексной таблицы на основе последней таблицы решения исходной задачи в случае, когда коэффициент $a_{i_0 j_0}$ при базисной переменной j_0 в ограничении безрезервного ресурса i_0 $(x_{m+i_0} = 0)$ получает приращение $\Delta_{i_0 j_0}$, примут следующий вид:

для строки i_0

$$x'^*_{i_0} = \frac{x^*_{i_0}}{1 + \alpha^*_{i,m+i_0} \Delta a_{i_0 j_0}},$$

$$\alpha'^*_{i_0 j} = \frac{\alpha^*_{i_0 j}}{1 + \alpha^*_{i_0,m+j_0} \Delta a_{i_0 j_0}};$$

для элементов любой строки $i \neq i_0$ и любого столбца $j \neq j_0$

$$x'^*_i = x^*_i - x^*_{i_0} \Delta a_{i_0 j_0} \frac{\alpha^*_{i,m+j_0}}{1 + \alpha^*_{i_0,m+j_0} \Delta a_{i_0 j_0}},$$

$$\alpha'^*_{ij} = \alpha^*_{ij} - \alpha_{i_0 j} \Delta a_{i_0 j_0} \frac{\alpha^*_{i,m+j_0}}{1 + \alpha^*_{i_0,m+j_0} \Delta a_{i_0 j_0}};$$

для элементов целевой функции

$$\Delta'^*_j = \Delta^*_j - \alpha^*_{i_0 j} \Delta a_{i_0 j_0} \frac{\Delta^*_{m+j_0}}{1 + \alpha^*_{i_0,m+j_0} \Delta a_{i_0 j_0}},$$

$$j \neq j_0 \quad (j = \overline{1, m+n}\,).$$

Для вычисления границ изменения $\Delta a_{i_0 j_0}$ нужно только найти x'^*_i $(i \in (\text{БП}))$ и Δ'^*_j. Эти границы находятся путем решения системы неравенств: $x'^*_i \geqslant 0$ $(i \in (\text{БП}))$, $\Delta'^*_j \geqslant 0$ $(j = \overline{1, m+n}\,)$.

3. ЭЛЕМЕНТЫ ТЕОРИИ МАТРИЧНЫХ ИГР

3.1. МАТРИЧНЫЕ ИГРЫ С НУЛЕВОЙ СУММОЙ

Теория игр занимается разработкой различного рода рекомендаций по принятию решений в условиях конфликтной ситуации. Формализуя конфликтные ситуации математически, их можно представить как игру двух, трех и более игроков, каждый из которых преследует цель максимизации своего выигрыша за счет другого игрока. Иногда теорию игр определяют как раздел математики, занимающийся выработкой оптимальных правил поведения для каждой стороны, участвующей в конфликтной ситуации. Совокупность правил, однозначно определяющих последовательность действий стороны в конкретной конфликтной ситуации, есть *стратегия.*

Под термином "*игра*" понимается совокупность предварительно оговоренных правил и условий, а термин "*партия*" связан с частичной возможной реализацией этих правил. Если n партнеров (игроков) $P_1, P_2, \ldots, P_n$ участвуют в данной игре, то основное содержание теории игр состоит в изучении следующей проблемы: как должен вести партию j-й партнер $(j = \overline{1, n})$ для достижения наиболее благоприятного для себя исхода?

В дальнейшем предполагается, что в конце партии каждый игрок P_j получает сумму v_j, называемую *выигрышем.* При этом подразумевается, что каждый игрок руководствуется лишь целью максимизации общей суммы выигрыша. Числа v_j $(j = \overline{1, n})$ могут быть положительными, отрицательными или равными нулю. Если $v_j > 0$, то это соответствует выигрышу j-го игрока, если $v_j < 0$, — проигрышу, при $v_j = 0$ — ничейный исход.

В большинстве случаев имеем игры с нулевой суммой, т. е. $v_1 + v_2 + \ldots + v_n = 0$. В этих играх сумма выигрыша переходит от одного партнера к другому, не поступая из внешних источников. Игра с нулевой суммой предусматривает, что сумма выигрышей всех игроков в каждой партии равна нулю. Примерами игры с нулевой суммой служат многие экономические задачи. В них общая сумма выигрыша перераспределяется между игроками, но не меняется. В противном случае имеем игру с ненулевой суммой.

Игры, в которых участвуют два игрока, называются *парными*, а игры с бо́льшим числом участников — *множественными*. Принятие игроком того или иного решения в процессе игры и его реализация называется *ходом*. Ходы могут быть *личные* и *случайные*. Если ход выбирается сознательно, — это личный ход, а если с помощью механизма случайного выбора, — случайный ход.

Шахматы являются игрой двух партнеров с конечным числом личных ходов. В дальнейшем мы будем рассматривать игры двух партнеров с нулевой суммой и конечным числом возможных ходов. Такие игры математически глубоко проработаны и вызывают наибольший интерес, поскольку чаще используются в практических приложениях.

В зависимости от количества стратегий игры делятся на *конечные* и *бесконечные*. Так, в конечной игре каждый из игроков имеет конечное число возможных стратегий. Если же хотя бы один из игроков имеет бесконечное число возможных стратегий, то игра называется бесконечной.

В зависимости от взаимоотношений игроков игры делятся на *кооперативные*, *коалиционные* и *бескоалиционные*. Если игроки не имеют права вступать в соглашения, то такая игра относится к бескоалиционным, если же игроки могут вступать в соглашения, создавать коалиции, — к коалиционным. Кооперативная игра — это такая игра, в которой заранее определены коалиции.

В зависимости от вида функции выигрышей игры подразделяются на *матричные*, *биматричные*, *непрерывные*, *выпуклые*, *сепарабельные* и т. д. Мы будем рассматривать матричные игры. Обратимся к примерам простейших матричных игр.

Пример 3.1. Первый игрок P_1 выбирает одну из двух сторон монеты. Второй игрок P_2, не зная выбора первого, также выбирает одну из сторон. После того как оба игрока произвели свой выбор и монета брошена, игрок P_2 платит 1 игроку P_1, если выбранные стороны монеты совпали, и -1 в противном случае. Здесь 1 соответствует выигрышу игроком P_1 одной единицы, а -1 соответствует проигрышу им одной единицы. В этом предположении мы говорим, что P_1 играет на максимум, а P_2 — на минимум.

Постановку задачи можно записать так:

Стратегии игроков		P_2	
		Орел	Решка
P_1	Орел	1	-1
	Решка	-1	1

Таким образом, условия игры определяются матрицей

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

строки которой соответствуют возможным стратегиям для P_1, а столбцы — возможным стратегиям для P_2. Как только P_1 выбирает строку и P_2 — столбец, партия заканчивается и выигрыш игрока P_1 равен числу, стоящему на пересечении выбранных строки и столбца.

Пример 3.2 ("игра в три пальца"). Игроки P_1 и P_2 одновременно и независимо друг от друга показывают 1, 2 или 3 пальца. Размер выигрыша определяется общим количеством показанных пальцев. При этом, если число пальцев четное, выигрывает игрок P_1, нечетное, — игрок P_2.

Такую игру двух игроков можно представить в виде матрицы

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 2 & -3 & 4 \\ -3 & 4 & -5 \\ 4 & -5 & 6 \end{bmatrix},$$

где индекс i элементов a_{ij} $(i = j = 1, 2, 3)$ означает количество пальцев игрока P_1, а индекс j — количество пальцев игрока P_2. Например, a_{13} означает, что одновременно и независимо друг от друга игрок P_1 показал 1 палец, а игрок P_2 — 3 пальца. Количество пальцев для элемента $a_{13} = 4$ указывает на выигрыш 4 единиц игроком P_1. Элемент $a_{32} = -5$ указывает на проигрыш 5 единиц игроком P_1 или выигрыш 5 единиц игроком P_2.

Мы рассмотрели примеры матричных игр 2-го и 3-го порядков. В общем случае матричная игра задается прямоугольной матрицей размерности $m \times n$. Номер i строки матрицы соответствует номеру стратегии A_i, применяемой игроком P_1. Номер j столбца соответствует стратегии B_j, применяемой игроком P_2. Описанная игра однозначно определяется матрицей

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [a_{ij}]_{m \times n}.$$

Каждый элемент a_{ij} матрицы является действительным числом и представляет собой сумму выигрыша, уплачиваемую игроком P_2 игроку P_1, если P_1 выбирает стратегию, соответствующую i-й строке, а P_2 выбирает стратегию, соответствующую j-му столбцу.

Матричную игру часто записывают в развернутой форме в виде табл. 3.1, называемой *платежной матрицей.*

Таблица 3.1

	B_1	$\dots$	B_j	$\dots$	B_n
A_1	a_{11}	$\dots$	a_{1j}	$\dots$	a_{1n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
A_i	a_{i1}	$\dots$	a_{ij}	$\dots$	a_{in}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
A_m	a_{m1}	$\dots$	a_{mj}	$\dots$	a_{mn}

Каждый игрок выбирает для себя наиболее выгодную стратегию. При этом первый игрок стремится выбрать такую стратегию, которая доставляет ему максимальный выигрыш, тогда второй игрок выбирает стратегию, приводящую его к минимальному проигрышу. В этой связи вводят понятия нижней и верхней чистой цены игры.

Нижней чистой ценой игры (*максимином*) называется число α, определяемое по формуле

$$\alpha = \max_i \min_j a_{ij}. \tag{3.1}$$

Верхней чистой ценой игры (*минимаксом*) называется число β, определяемое по формуле

$$\beta = \min_j \max_i a_{ij}. \tag{3.2}$$

Стратегии игроков, соответствующие максимину (минимаксу), называются *максиминными* (*минимаксными*).

Пример 3.3. Найти максиминную и минимаксную стратегии игроков в матричной игре

$$\begin{bmatrix} 2 & -3 & 4 & 5 \\ 3 & 7 & 8 & 4 \\ 5 & 1 & 3 & 7 \\ 4 & 6 & 2 & 9 \end{bmatrix}.$$

Решение. Данную игру представим в виде платежной матрицы (табл. 3.2).

Таблица 3.2

	B_1	B_2	B_3	B_4	α_i
A_1	2	−3	4	5	−3
A_2	3	7	8	4	3
A_3	5	1	3	7	1
A_4	4	6	2	9	2
β_j	5	7	8	9	

В соответствии с формулой (3.1) по каждой строке определяем наименьшее число, которое записывается в столбец α_i. Это означает, что, какой бы выбор по столбцам ни сделал игрок B, выигрыш игрока A, который свои стратегии выбирает по строкам, в худшем случае составит соответственно: −3, 3, 1, 2. Однако игроку A целесообразно выбрать такую стратегию (строку), для которой достигается максимальный выигрыш независимо от того, какой столбец выбрал игрок B, т. е.

$$\alpha = \max_i \min_j a_{ij} = \max_i \alpha_i = \max(-3,\ 3,\ 1,\ 2) = 3.$$

Максиминной стратегией игрока A является A_2.

Аналогично, пользуясь формулой (3.2), определяем минимаксную стратегию игрока B. Поскольку он выбирает стратегии по столбцам, то какие бы стратегии ни выбирал игрок A, в худшем случае игрок B может проиграть соответственно стратегиям B_1, B_2, B_3, B_4: 5, 7, 8, 9. Однако игрок B стремится минимизировать свой проигрыш, а потому выбирает стратегию, соответствующую минимальному из чисел 5, 7, 8, 9, т. е. минимаксу:

$$\beta = \min_j \max_i a_{ij} = \min_j \beta_j = \min(5,\ 7,\ 8,\ 9) = 5.$$

Из платежной матрицы видно, что минимаксной стратегией игрока B является B_1.

3.2. ЧИСТЫЕ И СМЕШАННЫЕ СТРАТЕГИИ И ИХ СВОЙСТВА

Различают стратегии *чистые* и *смешанные*. Чистая стратегия A_i $(i = \overline{1,m})$ первого игрока (чистая стратегия B_j $(j = \overline{1,n})$ второго игрока) — это возможный ход первого (второго) игрока, выбранный им с вероятностью, равной 1.

Если первый игрок имеет m стратегий, а второй — n стратегий, то для любой пары стратегий первого и второго игроков чистые стратегии можно представить в виде единичных векторов. Например, для пары стратегий A_1, B_2 чистые стратегии первого и второго игроков запишутся в виде: $\mathbf{p}_1 = (1;0;\ldots;0)$, $\mathbf{q}_2 = (0;1;0;\ldots;0)$. Для пары стратегий A_i, B_j чистые стратегии можно записать в виде:

$$\mathbf{p}_i = (0;\ \ldots\ ;\ 0;\ \underbrace{1;}_{i-\text{е место}}\ 0;\ \ldots\ ;\ 0),$$

$$\mathbf{q}_j = (0;\ \ldots\ ;\ 0;\ \underbrace{1;}_{j-\text{е место}}\ 0;\ \ldots\ ;\ 0).$$

Теорема 3.1. *В матричной игре нижняя чистая цена игры не превосходит верхней чистой цены игры, т. е. $\alpha \leqslant \beta$.*

Доказательство. По определению $\alpha_i = \min\limits_j a_{ij} \leqslant a_{ij}$. Аналогично $\beta_j = \max\limits_i a_{ij} \geqslant a_{ij}$. Объединив эти соотношения, получим

$$\alpha_i = \min_j a_{ij} \leqslant a_{ij} \leqslant \max_i a_{ij} = \beta_j.$$

Отсюда $\alpha_i \leqslant a_{ij} \leqslant \beta_j$ или $\alpha_i \leqslant \beta_j$ $(i = \overline{1,m};\ \ j = \overline{1,n})$. Это неравенство справедливо для любых i и j, следовательно, $\alpha \leqslant \beta$. $\triangle$

Если для чистых стратегий A_i, B_j игроков A и B соответственно имеет место равенство $\alpha = \beta$, то пару чистых стратегий (A_i, B_j) называют *седловой точкой матричной игры*, элемент a_{ij} матрицы, стоящий на пересечении i-й строки и j-го столбца, — *седловым элементом платежной матрицы*, а число $v = \alpha = \beta$ — *чистой ценой игры.*

Пример 3.4. Найти нижнюю и верхнюю чистые цены, установить наличие седловых точек матричной игры

$$\begin{bmatrix} 9 & 5 & 6 & 7 \\ 1 & 4 & 3 & 8 \\ 6 & 3 & 2 & -4 \end{bmatrix}.$$

Решение. Определим нижние и верхние чистые цены игры (табл. 3.3):

$$\alpha = \max \alpha_i = \max(5,\ 1,\ -4) = 5,$$
$$\beta = \min \beta_j = \min(9,\ 5,\ 6,\ 8) = 5,$$
$$v = \alpha = \beta = 5.$$

В данном случае имеем одну седловую точку $(A_1; B_2)$, а седловой элемент равен 5. Этот элемент является наименьшим в 1-й строке и наибольшим во 2-м столбце. Отклонение игрока A от максиминной стратегии A_1 ведет к уменьшению его выигрыша, а отклонение игрока B от минимаксной стратегии B_2 ведет к увеличению его проигрыша. Иными словами, если в матричной игре имеется седловой элемент, то наилучшими для игроков являются их минимаксные стратегии. И эти чистые стратегии, образующие седловую точку и выделяющие в матрице игры седловой элемент $a_{12} = 5$, есть оптимальные чистые стратегии A_1^* и B_2^* соответственно игроков A и B.

Таблица 3.3

	B_1	B_2	B_3	B_4	α_i
A_1	9	$\boxed{5}$	6	7	5
A_2	1	4	3	8	1
A_3	6	3	2	−4	−4
β_j	9	5	6	8	

Если же матричная игра не имеет седловой точки, то решение игры затрудняется. В этих играх $\alpha < \beta$. Применение минимаксных стратегий в таких играх приводит к тому, что для каждого из игроков выигрыш не превышает α, а проигрыш — не меньше β. Для каждого игрока возникает вопрос увеличения выигрыша (уменьшения проигрыша). Решение находят, применяя смешанные стратегии.

Смешанной стратегией первого (второго) игрока называется вектор $\mathbf{p} = (p_1; \dots; p_m)$, где $p_i \geqslant 0 \;\; (i = \overline{1,m})$ и $\sum\limits_{i=1}^{m} p_i = 1$ $(\mathbf{q} = (q_1; \dots; q_n)$, где $q_j \geqslant 0 \;\; (j = \overline{1,n})$ и $\sum\limits_{j=1}^{n} q_j = 1)$.

Вектор $\mathbf{p}$ $(\mathbf{q})$ означает вероятность применения i-й чистой стратегии первым игроком (j-й чистой стратегии вторым игроком).

Поскольку игроки выбирают свои чистые стратегии случайно и независимо друг от друга, игра имеет случайный характер и случайной становится величина выигрыша (проигрыша). В таком случае средняя величина выигрыша (проигрыша) — математическое ожидание — является функцией смешанных стратегий $\mathbf{p}, \mathbf{q}$:

$$f(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} p_i q_j.$$

Функция $f(\mathbf{p}, \mathbf{q})$ называется *платежной функцией* игры с матрицей $[a_{ij}]_{m \times n}$.

Стратегии $\mathbf{p}^* = (p_1^*; \ldots ; p_m^*)$, $\mathbf{q}^* = (q_1^*; \ldots ; q_n^*)$ называются *оптимальными*, если для произвольных стратегий $\mathbf{p} = (p_1; \ldots ; p_m)$, $\mathbf{q} = (q_1; \ldots ; q_n)$ выполняется условие

$$f(\mathbf{p}, \mathbf{q}^*) \leqslant f(\mathbf{p}^*, \mathbf{q}^*) \leqslant f(\mathbf{p}^*, \mathbf{q}). \tag{3.3}$$

Использование в игре оптимальных смешанных стратегий обеспечивает первому игроку выигрыш, не меньший, чем при использовании им любой другой стратегии $\mathbf{p}$, второму игроку — проигрыш, не больший, чем при использовании им любой другой стратегии $\mathbf{q}$.

Совокупность оптимальных стратегий и цены игры составляет *решение игры.*

Значение платежной функции при оптимальных стратегиях определяет цену игры v, т. е. $f(\mathbf{p}^*, \mathbf{q}^*) = v$.

Теорема 3.2. *В смешанных стратегиях любая конечная матричная игра имеет седловую точку.*

Пусть имеем матричную игру $[a_{ij}]_{m\times n}$ и некоторые смешанные оптимальные стратегии $\mathbf{p}^*, \mathbf{q}^*$ игроков A и B, обеспечивающие сумму выигрыша v. Вопрос поставим так: как проверить, что набор $(\mathbf{p}^*, \mathbf{q}^*, v)$ является решением игры? Для этого нужно проверить справедливость неравенства (3.3) для любых смешанных стратегий, среди которых и будут стратегии $\mathbf{p}^*, \mathbf{q}^*$. Однако различных смешанных стратегий, среди которых и оптимальные, имеем бесчисленное множество. И в таком случае проверить справедливость неравенства (3.3) невозможно. Поэтому рассмотрим следующую теорему, которая позволит ответить на поставленный выше вопрос.

Теорема 3.3. *Для того чтобы смешанные стратегии* $\mathbf{p}^* = (p_1^*; \ldots ; p_m^*)$ *и* $\mathbf{q}^* = (q_1^*; \ldots ; q_n^*)$ *были оптимальными для игроков* A *и* B *в игре с матрицей* $[a_{ij}]_{m\times n}$ *и выигрышем* v, *необходимо и достаточно выполнения неравенств:*

$$\sum_{i=1}^{m} a_{ij}p_i^* \geqslant v \quad (j = \overline{1, n}\,), \tag{3.4}$$

$$\sum_{j=1}^{n} a_{ij}q_j^* \leqslant v \quad (i = \overline{1, m}\,). \tag{3.5}$$

Доказательство. Пусть $\mathbf{p}^*$, $\mathbf{q}^*$ — оптимальные смешанные стратегии. Докажем, что для них выполняются соотношения (3.4) и (3.5). Воспользуемся определением оптимальных смешанных стратегий, для которых выполняется соотношение (3.3). Неравенство (3.4) получается из соотношения (3.3), если записать его в развернутой форме, а именно:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}p_i q_j^* \leqslant v \leqslant \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}p_i^* q_j. \tag{3.6}$$

В правую часть соотношения (3.6) подставим вектор $\mathbf{q}_j = (q_1; \ldots; q_{j-1}; q_j; q_{j+1}; \ldots; q_n) = (0; \ldots; 0; 1; 0; \ldots; 0)$. Получим

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}p_i^* q_j = \sum_{j=1}^{n} q_j \sum_{i=1}^{m} a_{ij}p_i^* = \sum_{i=1}^{m} a_{ij}p_i^* \geqslant v,$$

т. е. оптимальная стратегия $\mathbf{p}^*$ удовлетворяет неравенству (3.4).

Если вместо произвольного вектора $\mathbf{p}$ в левую часть соотношения (3.6) подставить вектор $\mathbf{p}_i = (p_1; \ldots; p_{i-1}; p_i; p_{i+1}; \ldots; p_m) = (0; \ldots; 0; 1; 0; \ldots; 0)$, то можно показать, что и оптимальная стратегия $\mathbf{q}^*$ удовлетворяет соотношению (3.5).

Итак, доказано условие необходимости, а именно: если стратегии $\mathbf{p}^*$ и $\mathbf{q}^*$ оптимальные, то они должны удовлетворять соотношениям (3.4) и (3.5).

Теперь докажем достаточность этого условия. Пусть выполняются неравенства (3.4), (3.5). Покажем, что $\mathbf{p}^*, \mathbf{q}^*$ — оптимальные стратегии. Для этого нужно показать выполнимость соотношения (3.6). С учетом соотношения (3.4) преобразуем правую часть, а с учетом соотношения (3.5) — левую часть соотношения (3.6).

Пусть $\mathbf{q}_j = (q_1; \ldots; q_n)$ — произвольный вектор, тогда

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}p_i^* q_j = \sum_{j=1}^{n} q_j \sum_{i=1}^{m} a_{ij}p_i^* \geqslant \sum_{j=1}^{n} q_j v = v\sum_{j=1}^{n} q_j = v \cdot 1 = v,$$

т. е. $\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_{ij}p_i^* q_j \geqslant v$.

Преобразуя левую часть соотношения (3.6) для произвольного вектора $\mathbf{p}_i = (p_1; \ldots; p_m)$, получаем

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}p_i q_j^* = \sum_{i=1}^{m} p_i \sum_{j=1}^{n} a_{ij}q_j^* \leqslant \sum_{i=1}^{m} p_i v = 1 \cdot v = v,$$

т. е. $\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_{ij}p_i q_j^* \leqslant v$.

Итак, доказано, что если выполняются соотношения (3.4), (3.5), то выполняется и соотношение (3.6), т. е. смешанные стратегии $\mathbf{p}^*$ и $\mathbf{q}^*$ — оптимальные.

Таким образом, для проверки того, что набор $(\mathbf{p}^*, \mathbf{q}^*, v)$ является решением матричной игры, достаточно проверить, удовлетворяют ли $\mathbf{p}^*, \mathbf{q}^*$ неравенствам (3.4) и (3.5) и уравнениям

$$\sum_{i=1}^{m} p_i = 1, \quad \sum_{j=1}^{n} q_j = 1. \ \triangle$$

На основании теоремы 3.3 можно сделать вывод: *если игрок A применяет оптимальную смешанную стратегию $\mathbf{p}^*$, а игрок B — любую чистую стратегию B_j, то выигрыш игрока A будет не меньше цены игры v.* Аналогично: *если игрок B использует оптимальную смешанную стратегию $\mathbf{q}^*$, а игрок A — любую чистую стратегию A_i, то проигрыш игрока B не превысит цены игры v.*

Чистые стратегии игрока, входящие в его оптимальную смешанную стратегию с вероятностями, отличными от нуля, называются *активными стратегиями игрока.* Рассмотрим теорему об активных стратегиях.

Теорема 3.4. *Если один из игроков придерживается своей оптимальной смешанной стратегии, то его выигрыш остается неизменным и равным цене игры независимо от того, какую стратегию применяет другой игрок, если только тот не выходит за пределы своих активных стратегий.*

Доказательство. Пусть в матричной игре $[a_{ij}]_{m \times n}$ имеем оптимальные стратегии игроков A и B соответственно $\mathbf{p}^*$ и $\mathbf{q}^*$. Цена игры равна v. При этом игрок A имеет r активных стратегий, а игрок B — k активных стратегий. Расположив активные стратегии для игроков первыми, будем иметь:

$\mathbf{p}^* = (p_1^*; \dots ; p_r^*; 0; \dots ; 0)$, $\mathbf{q}^* = (q_1^*; \dots ; q_k^*; 0; \dots ; 0)$, для которых

$$\sum_{i=1}^{r} p_i^* = 1, \quad \sum_{j=1}^{k} q_j^* = 1.$$

Пусть игрок A придерживается своей оптимальной стратегии $\mathbf{p}^*$, а игрок B — чистой стратегии, тогда, согласно теореме 3.3,

$$\sum_{i=1}^{r} a_{ij} p_i^* \geqslant v \quad (j = \overline{1, k}). \tag{3.7}$$

Если игроки A и B используют свои оптимальные стратегии, то выигрыш игрока A равен цене игры v, т. е.

$$v = \sum_{i=1}^{r} \sum_{j=1}^{k} a_{ij} p_i^* q_j^*.$$

Учитывая соотношение (3.7), получаем

$$v = \sum_{j=1}^{k} \sum_{i=1}^{r} a_{ij} p_i^* q_j^* = \sum_{j=1}^{k} q_j^* \sum_{i=1}^{r} a_{ij} p_i^* \geqslant \sum_{j=1}^{k} q_j^* v = v. \tag{3.8}$$

Соотношение (3.8) выполнимо лишь в случае, когда неравенства (3.7) превращаются в равенства. Отсюда можно сделать вывод, что для любой смешанной стратегии $\mathbf{q} = (q_1; \dots ; q_k; 0; \dots ; 0)$ выполняется равенство

$$\sum_{i=1}^{r} \sum_{j=1}^{k} a_{ij} p_i^* q_j = v,$$

что и доказывает теорему. $\triangle$

На основании данной теоремы решение матричной игры можно упростить, выявив при этом доминирование одних стратегий над другими. Так, рассматривая стратегии игрока A, сравниваем элементы строк s и t, а именно: a_{sj} с элементами a_{tj} для $j = \overline{1, n}$. Если $a_{sj} \geqslant a_{tj}$ $(j = \overline{1, n})$, то выигрыш игрока A при стратегии A_s будет больше, чем при стратегии A_t. В этом случае стратегия A_s доминирует над стратегией

A_t. Стратегию A_s называют *доминирующей*, а стратегию A_t — *доминируемой*.

Поскольку игрок B заинтересован в минимизации проигрыша, доминирующим будет столбец с наименьшими элементами. Например, сравниваем элементы r-го и l-го столбцов. Если все элементы $a_{ir} \geqslant a_{il}$ $(i = \overline{1,m})$, то игроку B свой выбор выгодно сделать по l-му столбцу. В этом случае стратегия B_l игрока B доминирует над стратегией B_r. Стратегия B_l называется *доминирующей*, а стратегия B_r — *доминируемой*.

Если в матричной игре имеем строки (столбцы) с одними и теми же элементами, то строки (столбцы), а соответственно и стратегии игроков A и B называются *дублирующими*.

В матричной игре доминируемые и дублирующие строки (столбцы) можно опускать, что не влияет на решение игры.

Теорема 3.5. *Оптимальные смешанные стратегии* $\mathbf{p}^*$ *и* $\mathbf{q}^*$ *соответственно игроков A и B в матричной игре* $[a_{ij}]_{m\times n}$ *с ценой v будут оптимальными и в матричной игре* $[ba_{ij}+c]_{m\times n}$ *с ценой $v' = bv + c$, где $b > 0$.*

Доказательство. На основании теоремы 3.3 для оптимальной смешанной стратегии $\mathbf{p}^*$ игрока A и для любой чистой стратегии B_j игрока B имеем

$$\sum_{i=1}^{m} a_{ij}p_i^* \geqslant v \quad (j = \overline{1,n}). \tag{3.9}$$

Умножим обе части неравенства (3.9) на некоторое положительное число $b > 0$ и к обеим частям полученного неравенства прибавим произведение $c\sum\limits_{i=1}^{m} p_i^*$. Получим

$$\sum_{i=1}^{m} ba_{ij}p_i^* + c\sum_{i=1}^{m} p_i^* \geqslant bv + c\sum_{i=1}^{m} p_i^* \quad (j = \overline{1,n}). \tag{3.10}$$

Так как $\sum\limits_{i=1}^{m} p_i^* = 1$, соотношение (3.10) примет вид

$$\sum_{i=1}^{m} (ba_{ij}+c)p_i^* \geqslant bv + c$$

или

$$\sum_{i=1}^{m}(ba_{ij}+c)p_i^* \geqslant v' \quad (j=\overline{1,n}),$$

где $v' = bv+c$. Теорема доказана для оптимальной смешанной стратегии $\mathbf{p}^*$ игрока A. $\triangle$

Аналогично доказывается теорема и для оптимальной смешанной стратегии игрока B.

На основании теоремы 3.5 платежную матрицу, имеющую отрицательные числа, можно преобразовать в матрицу с положительными числами.

Пример 3.5. Выполнить всевозможные упрощения матричной игры

$$\begin{bmatrix} 4 & -2 & 5 & 1 & 2 & 7 \\ 1 & 2 & 4 & 3 & 0 & 10 \\ 3 & 5 & 6 & 7 & 1 & 9 \\ 1 & 2 & 4 & 3 & 0 & 10 \\ 2 & 1 & 3 & 6 & 5 & 4 \end{bmatrix}.$$

Решение. Поскольку соответствующие элементы второй и четвертой строк матрицы игры равны, т. е. имеем две дублирующие строки, опустим, например, четвертую строку:

$$\begin{bmatrix} 4 & -2 & 5 & 1 & 2 & 7 \\ 1 & 2 & 4 & 3 & 0 & 10 \\ 3 & 5 & 6 & 7 & 1 & 9 \\ 2 & 1 & 3 & 6 & 5 & 4 \end{bmatrix}.$$

Сравним соответствующие элементы столбцов. Элементы первого столбца доминируют над элементами третьего и шестого столбцов, а элементы второго столбца доминируют над соответствующими элементами четвертого столбца. Игроку B невыгодно применять стратегии B_3, B_4 и B_6. Опускаем третий, четвертый и шестой столбцы и получаем матрицу

$$\begin{bmatrix} 4 & -2 & 2 \\ 1 & 2 & 0 \\ 3 & 5 & 1 \\ 2 & 1 & 5 \end{bmatrix}.$$

Элементы второй строки меньше соответствующих элементов третьей строки. Следовательно, игроку A невыгодна стратегия A_2. Опуская вторую строку, получаем упрощенную матрицу

$$\begin{bmatrix} 4 & -2 & 2 \\ 3 & 5 & 1 \\ 2 & 1 & 5 \end{bmatrix}.$$

Если требуется получить матрицу с положительными элементами, то достаточно прибавить к ее элементам, например, число 2.

3.3. ПРИВЕДЕНИЕ МАТРИЧНОЙ ИГРЫ К ЗАДАЧЕ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Пусть имеем игру размерности $m \times n$ с матрицей

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Обозначим через $\mathbf{p}^* = (p_1; \dots; p_m)$, $\mathbf{q}^* = (q_1; \dots; q_n)$ оптимальные смешанные стратегии игроков A и B. Стратегия $\mathbf{p}^*$ игрока A гарантирует ему выигрыш не меньше v, независимо от выбора стратегии B_j игроком B. Это можно записать так:

$$\left.\begin{array}{l} a_{11}p_1 + a_{21}p_2 + \ldots + a_{m1}p_m \geqslant v, \\ a_{12}p_1 + a_{22}p_2 + \ldots + a_{m2}p_m \geqslant v, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}p_1 + a_{2n}p_2 + \ldots + a_{mn}p_m \geqslant v, \end{array}\right\} \qquad (3.11)$$

где $p_1 + p_2 + \ldots + p_m = 1$; $p_i \geqslant 0$ $(i = \overline{1, m})$.

Аналогично стратегия $\mathbf{q}^*$ игрока B гарантирует ему проигрыш не больше v, независимо от выбора стратегии A_i игроком A, т. е.

$$\left.\begin{array}{l} a_{11}q_1 + a_{12}q_2 + \ldots + a_{1n}q_n \leqslant v, \\ a_{21}q_1 + a_{22}q_2 + \ldots + a_{2n}q_n \leqslant v, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}q_1 + a_{m2}q_2 + \ldots + a_{mn}q_n \leqslant v, \end{array}\right\} \qquad (3.12)$$

где $q_1 + q_2 + \ldots + q_n = 1$; $q_j \geqslant 0$ $(j = \overline{1, n})$.

Поскольку элементы платежной матрицы на основании теоремы 3.5 всегда можно сделать положительными, то и цена игры $v > 0$.

Преобразуем системы (3.11) и (3.12), разделив обе части каждого неравенства на положительное число v, и введем новые обозначения: $p_i/v = x_i$, $q_j/v = y_j$ $(i = \overline{1, m};\ j = \overline{1, n})$.

Получим:

$$\left.\begin{array}{l} a_{11}x_1 + a_{21}x_2 + \ldots + a_{m1}x_m \geqslant 1, \\ a_{12}x_1 + a_{22}x_2 + \ldots + a_{m2}x_m \geqslant 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{mn}x_m \geqslant 1, \end{array}\right\} \qquad (3.13)$$

где

$$x_1 + x_2 + \ldots + x_m = 1/v, \quad x_i \geqslant 0 \quad (i = \overline{1,m}), \qquad (3.14)$$

и

$$\left.\begin{array}{l} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n \leqslant 1, \\ a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n \leqslant 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}y_1 + a_{m2}y_2 + \ldots + a_{mn}y_n \leqslant 1, \end{array}\right\} \qquad (3.15)$$

где

$$y_1 + y_2 + \ldots + y_n = 1/v, \quad y_j \geqslant 0 \quad (j = \overline{1,n}). \qquad (3.16)$$

Так как игрок A стремится максимизировать цену игры v, то обратная величина $1/v$ будет минимизироваться, поэтому оптимальная стратегия игрока A определится из задачи линейного программирования следующего вида: найти минимальное значение функции $f(\mathbf{x}) = x_1 + x_2 + \ldots + x_m$ при ограничениях (3.13), (3.14).

Оптимальная смешанная стратегия игрока B определится решением задачи следующего вида: найти максимальное значение функции

$$\widetilde{f}(\mathbf{y}) = y_1 + y_2 + \ldots + y_n$$

при ограничениях (3.15), (3.16).

Решив пару двойственных задач графическим (для случая двух переменных) или симплексным методом, далее определим:

$$v = \frac{1}{\sum\limits_{i=1}^{m} x_i^*}, \quad p_i = \frac{x_i^*}{\sum\limits_{i=1}^{m} x_i^*}, \quad q_j = \frac{y_j^*}{\sum\limits_{j=1}^{n} y_j^*} \quad (i = \overline{1,m};\ j = \overline{1,n}).$$

Проиллюстрируем решение матричной игры сведéнием ее к ЗЛП.

Пример 3.6. Два сельскохозяйственных предприятия A и B выделяют денежные средства на строительство трех объектов. С учетом особенностей вкладов и местных условий прибыль предприятия A в зависимости от объема финансирования выражается элементами матрицы

$$\begin{bmatrix} 50 & 15 & 20 \\ 25 & 40 & 30 \\ 10 & 30 & 60 \end{bmatrix}.$$

Будем предполагать, что убыток предприятия B при этом равен прибыли предприятия A. Требуется найти оптимальные стратегии предприятий A и B.

Решение. Обозначим чистые стратегии предприятий A и B через A_1, A_2, A_3 и B_1, B_2, B_3 соответственно. Предположим, что предприятие A располагает общей суммой a тыс. ден. ед., отпускаемой на строительство трех объектов. Аналогично и предприятие B имеет сумму в b тыс. ден. ед., отпускаемую на строительство тех же трех объектов. Тогда чистая стратегия A_1 — это выделение a_1 тыс. ден. ед. предприятием A на строительство первого объекта; A_2 — чистая стратегия предприятия A, которое выделяет сумму a_2 тыс. ден. ед. на строительство второго объекта; A_3 — чистая стратегия предприятия A, которое выделяет сумму a_3 тыс. ден. ед. на строительство третьего объекта. Общая сумма средств, выделяемых на строительство трех объектов, $a = a_1 + a_2 + a_3$. Аналогично определяются чистые стратегии и для предприятия B.

Проверим игру на наличие седловой точки:

$$\alpha = \max_i \min_j a_{ij} = 25, \quad \beta = \min_j \max_i a_{ij} = 40, \quad \alpha \neq \beta,$$

поэтому решение игры определяем в смешанных стратегиях. Цена игры v заключена между нижней α и верхней β ценами, т. е. $25 \leqslant v \leqslant 40$. Составим ЗЛП для каждого игрока.

Для игрока A: найти минимальное значение функции

$$f = x_1 + x_2 + x_3$$

при ограничениях

$$\left.\begin{aligned} 50x_1 + 25x_2 + 10x_3 &\geqslant 1, \\ 15x_1 + 40x_2 + 30x_3 &\geqslant 1, \\ 20x_1 + 30x_2 + 60x_3 &\geqslant 1, \end{aligned}\right\}$$

$$x_i \geqslant 0 \quad (i = 1,\ 2,\ 3).$$

Для игрока B: найти максимальное значение функции

$$\tilde{f} = y_1 + y_2 + y_3$$

при ограничениях

$$\left.\begin{aligned} 50y_1 + 15y_2 + 20y_3 &\leqslant 1, \\ 25y_1 + 40y_2 + 30y_3 &\leqslant 1, \\ 10y_1 + 30y_2 + 60y_3 &\leqslant 1, \end{aligned}\right\}$$

$$y_j \geqslant 0 \quad (j = 1,\ 2,\ 3).$$

Вводя вспомогательные переменные $x_4 \geqslant 0$, $x_5 \geqslant 0$, $x_6 \geqslant$ для исходной задачи и $y_4 \geqslant 0$, $y_5 \geqslant 0$, $y_6 \geqslant 0$ для двойственной задачи, модели задач преобразуем к канонической форме. При этом вспомогательные переменные примем за базисные. Соответствие между переменными пары взаимно двойственных задач будет следующее:

$$\begin{array}{ccc}
\overbrace{\begin{array}{ccc} x_1 & x_2 & x_3 \\ \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_6 \end{array}}^{\text{Свободные}} & \qquad & \overbrace{\begin{array}{ccc} x_4 & x_5 & x_6 \\ \updownarrow & \updownarrow & \updownarrow \\ y_1 & y_2 & y_3 \end{array}}^{\text{Базисные}} \\
\text{Базисные} & & \text{Свободные}
\end{array}$$

Решим, например, двойственную ЗЛП, построенную для определения выигрыша предприятия B. Каноническая форма задачи имеет вид:

$$\max f = y_1 + y_2 + y_3;$$

$$\left.\begin{array}{l}
50y_1 + 15y_2 + 20y_3 + y_4 = 1, \\
25y_1 + 40y_2 + 30y_3 + y_5 = 1, \\
10y_1 + 30y_2 + 60y_3 + y_6 = 1.
\end{array}\right\}$$

Решая ее симплекс-методом, приходим к табл. 3.4, в которой содержится оптимальный план $\mathbf{y}^* = (y_1^*; y_2^*; y_3^*; y_4^*; y_5^*; y_6^*) = (0{,}0133;\ 0{,}0094;\ 0{,}0098;\ 0;\ 0;\ 0)$. При этом $\widetilde{f}(\mathbf{y}^*) = 0{,}0325$.

Таблица 3.4

БП	$c_Б$	$\mathbf{A}_0$	y_1	y_2	y_3	y_4	y_5	y_6
			1	1	1	0	0	0
y_1	1	0,0133	1	0	0	0,0234	−0,0047	−0,0055
y_2	1	0,0094	0	1	0	−0,0188	0,0438	−0,0156
y_3	1	0,0098	0	0	1	0,0056	−0,0211	0,0254
$f_j - c_j$		0,0325	0	0	0	0,0102	0,0180	0,0043

С учетом основной теоремы двойственности и соответствия между переменными оптимальный план исходной задачи запишется в виде $\mathbf{x}^* = (x_1^*; x_2^*; x_3^*; x_4^*; x_5^*; x_6^*) = (0{,}0102;\ 0{,}0180;\ 0{,}0043;\ 0;\ 0;\ 0)$ $f(\mathbf{x}^*) = 0{,}0325$.

По формулам

$$v = \frac{1}{f_{\max}}, \quad \frac{p_i}{v} = x_i, \quad \frac{q_j}{v} = y_j \quad (i = \overline{1,m};\ j = \overline{1,n})$$

получим цену игры $v = \frac{1}{0{,}0325} \approx 30{,}77$ и вероятности p_i^* и q_j^* для оптимальных смешанных стратегий соответственно предприятий A и B:

$$p_1^* = 0{,}0102 \cdot 30{,}77 = 0{,}314, \quad q_1^* = 0{,}0133 \cdot 30{,}77 = 0{,}409,$$
$$p_2^* = 0{,}0180 \cdot 30{,}77 = 0{,}554, \quad q_2^* = 0{,}0094 \cdot 30{,}77 = 0{,}289,$$
$$p_3^* = 0{,}0043 \cdot 30{,}77 = 0{,}132, \quad q_3^* = 0{,}0098 \cdot 30{,}77 = 0{,}302.$$

Таким образом, оптимальными смешанными стратегиями сельскохозяйственных предприятий A и B являются стратегии $\mathbf{p}^* = (0{,}314;\ 0{,}554;\ 0{,}132)$ и $\mathbf{q}^* = (0{,}409;\ 0{,}289;\ 0{,}302)$ соответственно при гарантированном получении предприятием A независимо от стратегий предприятия B прибыли не менее 30,77 тыс. ден. ед. Убыток предприятия B при этом составит не более 30,77 тыс. ден. ед.

Итак, из общей суммы средств a тыс. ден. ед., выделяемых предприятием A на строительство трех объектов, на долю первого объекта должно выделяться 31,4 %, второго — 55,4 и третьего — 13,2 % этой суммы. Аналогично распределяются средства b тыс. ден. ед. предприятием B. Так, на долю первого объекта приходится 40,9 %, второго — 28,9 и третьего — 30,2 % общей суммы. Такое распределение денежных средств предприятиями A и B по трем строящимся объектам позволит им получить максимальную прибыль 30,77 тыс. ден. ед.

3.4. ИГРЫ С ПРИРОДОЙ. КРИТЕРИИ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ

Управление производственными процессами осуществляется путем реализации последовательности принимаемых решений. Для этого необходима информация о состоянии объекта управления в условиях его работы. В случае отсутствия достаточно полной информации возникает неопределенность в принятии решения. Причины этого могут быть различными: невозможность получения информации к моменту принятия решения; слишком высокие затраты на получение информации; невозможность устранения неопределенности по причинам объективного характера и т. д.

Естественно, по мере совершенствования средств сбора информации, передачи и обработки ее неопределенность ситуации в момент принятия управленческих решений будет уменьшаться. Существование неустранимой неопределенности связано со случайным характером многих явлений. Например, случайный характер спроса на продукцию делает невозможным точное прогнозирование объема ее выпуска. Принятие решения в этом случае связано с риском. Или, например, прием партии товара для контроля на соответствие стандарту также связан с риском. Правда, неопределенность при контроле может быть устранена в случае контроля всего товара, выпускаемого для реализации. Однако это может оказаться слишком дорогостоящим мероприятием.

С целью уменьшения неблагоприятных последствий в каждом конкретном случае следует учитывать степень риска и имеющуюся информацию. И здесь *лицо, принимающее решение* (ЛПР), вступает в игровые отношения с некоторым абстрактным лицом, которое условно можно назвать "*природой*". Иными словами, ЛПР должно уметь находить управленческое решение, когда природа не выбирает сознательно свои оптимальные стратегии. Вместе с тем мы иногда располагаем некоторыми вероятностными характеристиками состояния природы. Такого рода ситуации принято называть *играми с природой.*

Любую хозяйственную деятельность человека можно рассматривать как игру с природой. В широком смысле под "природой" будем понимать совокупность неопределенных факторов, влияющих на эффективность принимаемых решений.

Задачей экономиста или ЛПР является принятие наилучшего управленческого решения в каждой конкретной ситуации. Качество принимаемого решения зависит от информированности ЛПР о ситуации, в которой принимается решение. В случае неопределенности неквалифицированный экономист отказывается принимать решение или принимает его без достаточного обоснования. Хороший экономист руководствуется правилом: "информация — это деньги". Умение использовать даже неполную информацию для обоснования принимаемых решений — это задача экономиста.

Безразличие природы к результату игры (выигрышу) и возможность получения экономистом (статистиком) или ЛПР дополнительной информации о ее состоянии отличают игру с природой от обычной матричной игры, в которой принимают участие два сознательных игрока.

Игры с природой представляют собой основную модель теории принятия решений в условиях частичной неопределенности.

Множество состояний природы обозначим через Π, отдельное состояние — Π_j, $\Pi_j \in \Pi$ $(j = \overline{1,n})$. Множество решений (стратегий) статистика обозначим через A, отдельное решение — A_i, $A_i \in A$ $(i = \overline{1,m})$.

Если на множествах состояний природы Π и решений статистика A потребуется определить распределение вероятностей, то необходимо производить эксперимент, целью которого будет нахождение распределения некоторой случайной переменной, зависящей от состояния природы.

Для i-го решения A_i статистика A и j-го состояния природы Π_j имеем некоторое число, обозначающее функцию потерь $L(A_i, \Pi_j)$, которая, как правило, является случайной переменной.

Во взаимоотношениях с природой статистик может использовать любые из стратегий $A_1, \ldots, A_m$ в зависимости от состояний Π_j природы. Имея ряд стратегий $A_1, \ldots, A_m$, статистик должен руководствоваться некоторым правилом поведения, с помощью которого он определяет, какую стратегию $A_i \in A$ ему выбрать. Иными словами, статистик отыскивает оптимальное поведение, которое и будет его оптимальной стратегией. При этом он может пользоваться как чистыми, так и смешанными стратегиями.

Чтобы выразить в количественной форме упомянутое выше некоторое правило поведения статистика, которым он должен руководствоваться, предположим, что есть возможность численно оценить величиной a_{ij} эффективность каждой комбинации (A_i, Π_j), иначе говоря, качество решения A_i. Тем самым будет определена так называемая платежная матрица игры с природой

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix},$$

на основе которой в дальнейшем и будут сформулированы "правила поведения" — критерии выбора оптимальной стратегии статистика.

Элемент a_{ij} назовем выигрышем статистика, если он использует стратегию A_i при состоянии природы Π_j.

Решение игры с природой несколько отличается от решения обычной матричной игры, где оба игрока ведут игру сознательно. Отличие состоит прежде всего в упрощении игры. Выявление дублирующих и доминируемых стратегий производится только для стратегий статистика. Стратегии природы нельзя опускать, поскольку она не имеет "умысла" навредить статистику, более того, она может реализовать состояния, заведомо выгодные статистику. Иногда при решении игры с природой используется *матрица рисков.* Элементы r_{ij} матрицы рисков равны разности между максимально возможным выигрышем и тем выигрышем, который статистик получит в тех же условиях Π_j, применяя стратегию A_i, т. е. $r_{ij} = \beta_j - a_{ij}$, где $\beta_j = \max\limits_i a_{ij}$.

Оптимальную стратегию статистика можно определить, используя ряд критериев. Так, при известном распределении вероятностей различных состояний Π_j природы пользуются *критерием Байеса.* Показателем в этом критерии служит либо величина среднего выигрыша, либо величина среднего риска.

Платежную матрицу $[a_{ij}]_{m\times n}$ представим в виде табл. 3.5.

Таблица 3.5

Стратегия статистика A_i	Состояния природы Π_j				Средний выигрыш $\overline{a}_i$
	Π_1	Π_2	...	Π_n	
A_1	a_{11}	a_{12}	...	a_{1n}	$\overline{a}_1$
A_2	a_{21}	a_{22}	...	a_{2n}	$\overline{a}_2$
...	...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}	$\overline{a}_m$
q_j	q_1	q_2	...	q_n	

По критерию Байеса за оптимальную принимается та чистая стратегия A_i, при которой максимизируется средний выигрыш $\overline{a}_i$ статистика, т. е. обеспечивается $\overline{a} = \max\limits_i \overline{a}_i$, где $\overline{a}_i = \sum\limits_{j=1}^{n} a_{ij}q_j \quad (i = \overline{1,m})$.

Матрицу рисков представим в виде табл. 3.6.

За оптимальную стратегию статистика принимается чистая стратегия A_i, при которой минимизируется средний риск, т. е. обеспечивается $\overline{r} = \min\limits_i \overline{r}_i$, где $\overline{r}_i = \sum\limits_{j=1}^{n} r_{ij}q_j \quad (i = \overline{1,m})$.

В случае, когда вероятности состояний природы правдоподобны, для их оценки используют *принцип недостаточного* *основания Лапласа*, согласно которому все состояния природы полагаются равновероятными, т. е. $q_1 = q_2 = \ldots = q_n = 1/n$. Оптимальной считается стратегия, обеспечивающая максимум среднего выигрыша.

Таблица 3.6

Стратегия статистика A_i	Состояния природы $П_j$				Средний риск $\bar{r}_i$
	$П_1$	$П_2$	...	$П_n$	
A_1	r_{11}	r_{12}	...	r_{1n}	$\bar{r}_1$
A_2	r_{21}	r_{22}	...	r_{2n}	$\bar{r}_2$
...	.....	.....	.....	.....	...
A_m	r_{m1}	r_{m2}	...	r_{mn}	$\bar{r}_m$
q_j	q_1	q_2	...	q_n	

Если вероятности состояний природы неизвестны, то для решения игр с природой — выбора оптимальной стратегии статистика — можно использовать несколько критериев.

Максиминный критерий Вальда совпадает с критерием выбора максиминной стратегии, позволяющей получать нижнюю чистую цену α в парной игре с нулевой суммой. По критерию Вальда за оптимальную принимается чистая стратегия, которая в наихудших условиях гарантирует максимальный выигрыш, т. е.

$$\alpha = \max_i \min_j a_{ij}.$$

Критерий минимального риска Сэвиджа рекомендует выбирать в качестве оптимальной стратегии ту, при которой величина максимального риска минимизируется в наихудших условиях, т. е. обеспечивается $\min_i \max_j r_{ij}$.

Критерии Вальда и Сэвиджа ориентируют статистика на самые неблагоприятные состояния природы, т. е. эти критерии выражают пессимистическую оценку ситуации.

Критерий Гурвица является критерием пессимизма-оптимизма. За оптимальную принимается та стратегия, для которой выполняется соотношение

$$\max_i \left(\lambda \min_j a_{ij} + (1-\lambda) \max_j a_{ij} \right),$$

где $0 \leqslant \lambda \leqslant 1$. При $\lambda = 0$ имеем критерий крайнего оптимизма, а при $\lambda = 1$ — критерий пессимизма Вальда. Если $0 < \lambda < 1$, то имеем нечто среднее. При желании подстраховаться в данной ситуации λ принимают близким к единице. В общем случае число λ выбирают исходя из опыта или субъективных соображений.

Решение игры с природой по рассмотренным критериям позволяет более обоснованно принимать ту стратегию, которая гарантирует статистику больший выигрыш по сравнению с выигрышем, принимаемым статистиком интуитивно или исходя из опыта.

Пример 3.7. В соответствии со спросом на продукцию q-й номенклатуры в городе планируется построить предприятие по производству этой продукции. Неопределенность спроса в период t приводит к тому, что необходимо рассчитать объем выпускаемой продукции V_q, который должен быть не меньше уровня спроса S_q, чтобы не потерять потенциально возможный доход от реализации продукции, а также не должен превышать уровень спроса, так как предприятие будет нести убытки, связанные в основном с уценкой. Предположим, что в течение года (по кварталам) спрос на продукцию q-й номенклатуры выражается величинами 10, 20, 30, 40 тыс. шт. В таком случае и планирующий орган предприятия может принять одно из следующих решений: построить предприятие, которое могло бы удовлетворить спрос потребителей в 10, 20, 30, 40 тыс. шт. q-й продукции. Работа подобных предприятий показывает, что предприятие терпит издержки от нереализованной единицы q-й продукции 5 ден. ед., а доход от реализации единицы продукции составляет 15 ден. ед. Функцию платежей можно записать в виде кусочно-линейной функции потерь:

$$L(S_q, V_q) = \begin{cases} K'_q(V_q - S_q), & V_q \geqslant S_q, \\ K''_q(S_q - V_q), & V_q < S_q. \end{cases}$$

Требуется: 1) придать описанной ситуации игровую схему, установить характер игры и выявить ее участников; 2) вычислить элементы платежной матрицы и составить ее; 3) дать обоснованные рекомендации планирующему органу на строительство предприятия, которое могло бы обеспечить спрос потребителей на q-ю продукцию.

При изучении работы аналогичных предприятий планирующий орган располагает некоторой дополнительной информацией, снижающей неопределенность ситуации: известны вероятности спроса на данную продукцию по кварталам года: 0,3; 0,2; 0,4; 0,1; спрос на продукцию в каждом квартале равновероятен; о вероятностях спроса на указанную продукцию по кварталам ничего определенного сказать нельзя.

Решение. В качестве статистика выступает планирующий орган, который может принять одно из следующих решений: построить предприятие, способное удовлетворить спрос потребителей в 10 тыс. ед. продукции (стратегия A_1); построить предприятие мощностью в 20 тыс. ед.

продукции (стратегия A_2); построить предприятие мощностью в 30 тыс. ед. продукции (стратегия A_3); построить предприятие мощностью в 40 тыс. ед. продукции (стратегия A_4).

Второй играющей стороной — природой — будем считать совокупность объективных внешних условий, в которых формируется спрос потребителей. Спрос по кварталам года различен: в первом квартале — 10 тыс. ед. — будет означать состояние $П_1$; спрос на продукцию во втором квартале в объеме 20 тыс. ед. — состояние $П_2$; спрос на продукцию в третьем квартале в объеме 30 тыс. ед. — состояние $П_3$; спрос на продукцию в четвертом квартале в объеме 40 тыс. ед. — состояние $П_4$.

Итак, описанная ситуация представляет собой игру с природой.

Рассчитаем элементы платежной матрицы (табл. 3.7).

Таблица 3.7

Мощность предприятия, тыс. ед.	Стратегии статистика A_i	Спрос потребителей по кварталам года, тыс. ед.			
		10	20	30	40
		Состояние спроса $П_j$			
		$П_1$ (10)	$П_2$ (20)	$П_3$ (30)	$П_4$ (40)
10	A_1	150	150	150	150
20	A_2	100	300	300	300
30	A_3	50	250	450	450
40	A_4	0	200	400	600
	q_j	0,3	0,2	0,4	0,1

Так, в ситуации $(A_1, П_1)$ элемент a_{11} вычисляется следующим образом. Плановый орган принимает решение построить предприятие мощностью в 10 тыс. ед., что и соответствует состоянию спроса в 10 тыс. ед. Доход от производства 10 тыс. ед. продукции $a_{11} = 10 \cdot 15 = 150$ тыс. ден. ед.

Элемент a_{12} в ситуации $(A_1, П_2)$ рассчитываем так. Предприятие строится в расчете на выпуск 10 тыс. ед. продукции, а спрос на нее составляет 20 тыс. ед. Если бы предприятие могло обеспечить этот спрос, то доход составил бы $20 \cdot 15 = 300$ тыс. ден. ед. Однако спрос удовлетворяется лишь на 10 тыс. ед. продукции, следовательно, предприятие недополучит доход $10 \cdot 15 = 150$ тыс. ден. ед. Элемент $a_{12} = 20 \cdot 15 - 10 \cdot 15 = 150$.

Аналогично определяются и другие элементы табл. 3.7, например элемент a_{31} для ситуации $(A_3, П_1)$. Предприятие строится в расчете на выпуск 30 тыс. ед. продукции, а спрос на нее составляет 10 тыс. ед., тогда доход предприятия от реализации 10 тыс. ед. составит $10 \cdot 15 = 150$ тыс. ден. ед., а от нереализованной продукции (20 тыс. ед.) предприятие терпит издержки $20 \cdot 5 = 100$ тыс. ден. ед., доход предприятия в ситуации $(A_3, П_1)$ составит $a_{31} = 10 \cdot 15 - 20 \cdot 5 = 50$ тыс. ден. ед.

Аналогично рассчитываем и остальные элементы платежной матрицы (табл. 3.7).

Вычисляем средние выигрыши:

$$\overline{a}_1 = 150 \cdot 0,3 + 150 \cdot 0,2 + 150 \cdot 0,4 + 150 \cdot 0,1 = 150.$$

Аналогично $\overline{a}_2 = 240$, $\overline{a}_3 = 290$, $\overline{a}_4 = 260$.

Оптимальной стратегией по Байесу является A_3, поскольку ей соответствует максимальная средняя прибыль

$$\overline{a}_3 = \max(150,\ 240,\ 290,\ 260) = 290 \text{ (тыс. ден. ед.)}.$$

По критерию Лапласа, когда $q_1 = q_2 = q_3 = q_4 = 1/4$, средние выигрыши равны:

$$\overline{\alpha}_1 = \frac{1}{4}(150 + 150 + 150 + 150) = 150, \quad \overline{\alpha}_2 = 250, \quad \overline{\alpha}_3 = 300, \quad \overline{\alpha}_4 = 300.$$

Оптимальными стратегиями по Лапласу являются A_3 и A_4, так как им соответствует максимальная прибыль, равная 300 тыс. ден. ед.

По критерию Вальда оптимальной является стратегия A_1, для которой прибыль достигает наибольшей величины, равной 150 тыс. ден. ед. В самом деле,

$$\alpha = \max_i \min_j a_{ij} = \max(150,\ 100,\ 50,\ 0) = 150 \text{ (тыс. ден. ед.)}.$$

Построим матрицу рисков (табл. 3.8).

Таблица 3.8

Мощность предприятия, тыс. ед.	Стратегия статистика A_i	Спрос потребителей по кварталам года, тыс. ед.				$\max_j r_{ij}$
		10	20	30	40	
		Состояние спроса Π_j				
		Π_1	Π_2	Π_3	Π_4	
10	A_1	0	150	300	450	450
20	A_2	50	0	150	300	300
30	A_3	100	50	0	150	150
40	A_4	150	100	50	0	150

По критерию Сэвиджа оптимальными являются стратегии A_3 и A_4, для которых в наихудших условиях величина r риска принимает наименьшее значение, равное 150 тыс. ден. ед. В самом деле,

$$r = \min_i \max_j r_{ij} = \min(450,\ 300,\ 150,\ 150) = 150.$$

Таким образом, в результате решения игры с природой по различным критериям чаще других рекомендовалась стратегия A_3. Следовательно, нужно строить предприятие мощностью в 30 тыс. ед. продукции. Прибыль при этом, если вероятности спроса известны, составит 290 тыс. ден. ед., при равновероятных условиях спроса — 300 тыс. ден. ед.

4. ПРОГРАММИРОВАНИЕ НА СЕТЯХ

4.1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ГРАФОВ

Теория графов обладает широким разнообразием возможностей ее применения в различных областях знаний. В частности, она является эффективным аппаратом формализации задач экономической и планово-производственной практики. В последнее время теория графов нашла важные применения в автоматизации управления производством, в календарном и сетевом планировании, при оптимизации размещения производства, в теории массового обслуживания, при рационализации схем перевозок продукции, при наиболее компактной записи и обработке экономической информации и др. Основным объектом этой теории является *граф*.

Чтобы составить наглядное представление о графе, достаточно вообразить некоторое множество точек плоскости или пространства и множество отрезков кривых или прямых линий, соединяющих все или некоторые из этих точек. Формально же граф G определяется заданием двух множеств X и U и обозначается $G = \{X, U\}$. Элементы множества X называют *вершинами*. Будем их обозначать буквами $x_1, x_2, \ldots, x_n$. Вершины изображают точками плоскости или пространства. Элементами множества U являются пары связанных между собой элементов множества X. Их изображают отрезками кривых или прямых линий. Обозначать их будем буквами $u_1, u_2, \ldots, u_m$. Взаимное расположение, форма и длины упомянутых отрезков значения не имеют. Важно лишь, что они соединяют две данные вершины множества X. Если в паре вершин x_i и x_j указано направление связи, т. е. какая из них является первой, то соединяющий их отрезок u_k называется *дугой*; если же ориентация не указана, — *ребром*; а вершины, определяющие дугу (ребро) u_k, называют *концевыми вершинами дуги (ребра)* u_k. Если концевые вершины совпадают, то дугу (ребро) называют *петлей*. На графе могут существовать дуги (ребра) с одинаковыми концевыми вершинами. Будем называть такие дуги (ребра) *параллельными*. Говорят, что дуга (ребро) инцидентна своим концевым вершинам, и обратно.

Если в графе G все элементы множества U изображаются дугами, то граф называется *ориентированным* (*орграфом*) (рис. 4.1, *а*), если ребрами, — *неориентированным* (рис. 4.1, *б*). Иногда рассматривают *смешанные графы*, состоящие из

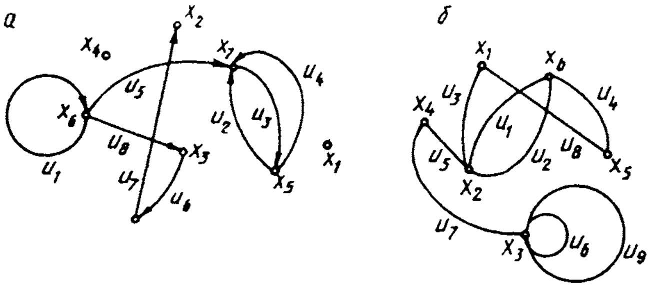


Рис. 4.1

дуг и ребер. Не всегда точки пересечения дуг (ребер) являются вершинами графа (см. рис. 4.1). Геометрическое изображение графа часто бывает весьма наглядным. Этим и пользуются на практике. Примерами графов могут служить схемы железных или шоссейных дорог, схемы связи поставщиков и потребителей, структурные формулы молекул и т. д. В приложениях обычно используются графы, в которых множества X и U состоят из конечного числа элементов. Такие графы называют *конечными*.

Поскольку при изображении графа его вершины можно располагать произвольно и по своему усмотрению выбирать форму соединяющих их линий, то один и тот же граф можно представить в различных видах. Так, на рис. 4.2 изображен по существу один и тот же граф, так как во всех вариантах (*а*, *б*, *в*) содержится одна и та же информация. О таких графах говорят, что они *изоморфны*. Два графа G и G' изоморфны, если между множествами их вершин существует такое взаимно однозначное соответствие, при котором в одном из графов отрезками соединены вершины в том и только в том случае, если в другом графе отрезками соединены соответствующие вершины. Понятно, что для орграфов ориентация дуг должна быть одинаковой.

Вершины в графе называют *смежными*, если они различны и существует дуга (ребро), соединяющая эти вершины. Если две дуги (ребра) имеют общую концевую вершину, они называются *смежными*. Вершины в графе могут отличаться друг от друга количеством дуг (ребер), которым они инцидентны (принадлежат). *Степенью* $P(x_i)$ *вершины* x_i называется число дуг (ребер) графа G, инцидентных данной вершине. Так, на рис. 4.1, *б* степень вершины x_2 равна 4, т. е. $P(x_2) = 4$. Вершина степени 0 называется *изолированной*. На рис. 4.1, *а* вершины x_1 и x_4 изолированные.

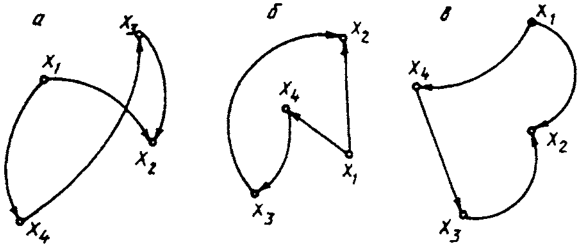


Рис. 4.2

Для орграфа вводятся понятия *полустепени захода* $P^+(x_i)$ *вершины* x_i — количество дуг, заходящих в x_i, и *полустепени исхода* $P^-(x_i)$ — количество дуг, исходящих из x_i. Понятно, что $P^+(x_i) + P^-(x_i) = P(x_i)$. На рис. 4.1, *а* $P^+(x_7) = 3,\ P^-(x_7) = 1$.

Граф называют *простым*, если он не содержит петель и параллельных дуг (ребер). Простой граф, в котором каждая пара вершин смежна, называется *полным*. Граф, содержащий хотя бы две параллельные дуги (ребра), называется *мультиграфом*. На рис. 4.1 изображены мультиграфы, так как граф на рис. 4.1, *а* содержит параллельные дуги u_2, u_3, u_4; в графе на рис. 4.1, *б* ребра u_1 и u_2 параллельны.

Путем в орграфе называется последовательность дуг, в которой конец каждой предыдущей дуги совпадает с началом следующей. Путь, проходящий через все вершины, и притом только по одному разу, называется *гамильтоновым* (рис. 4.3, $x_1x_2x_3x_4x_5$). Путь, содержащий все дуги графа, и притом только по одному разу, называется *эйлеровым*. Конечный путь, у которого начальная вершина совпадает с конечной, называется *контуром* (рис. 4.3, $x_1x_2x_3x_5x_1$). Контур, проходящий через каждую вершину графа только по одному разу (за исключением начальной и конечной вершин), называется *гамильтоновым*.

В неориентированном графе последовательность ребер, в которой каждые два соседних ребра смежны, называется *путем*, а конечный путь, у которого начальная и конечная вершины совпадают, — *циклом*.

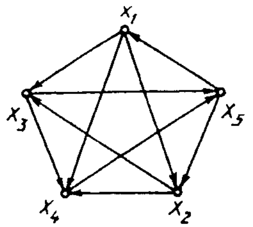


Рис. 4.3

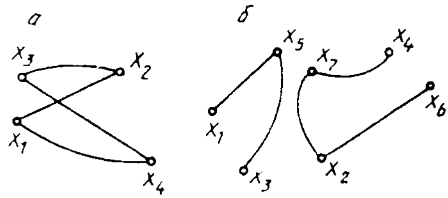


Рис. 4.4

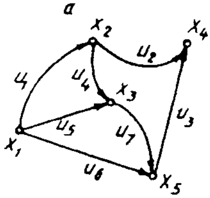
Важным понятием в теории графов является *связность*. Неориентированный граф называется *связным*, если любые две его вершины можно соединить путем (рис. 4.4, *а*), и *несвязным* в противном случае (рис. 4.4, *б*). Для установления связности орграфа ориентацию его дуг принимать в расчет не следует. Для орграфа существует еще понятие *сильной связности*. Говорят, что орграф сильно связный, если между любыми двумя его вершинами существует хотя бы один путь.

В различных приложениях теории графов дугам (ребрам) графов, моделирующим реальные процессы, обычно сопоставляют какие-либо числовые характеристики. Например, если дугами изображаются транспортные магистрали, то числовой характеристикой дуги может быть пропускная способность соответствующей магистрали; если же дугами изображаются работы некоторого комплекса, то числовая характеристика может означать время выполнения работы, количество некоторого ресурса, необходимого для ее выполнения, и т.д. В подобных случаях говорят, что дугам графа приписаны определенные веса, а граф G с весами на дугах называют *взвешенным*.

4.2. МАТРИЧНЫЕ СПОСОБЫ ЗАДАНИЯ ГРАФОВ. УПОРЯДОЧЕНИЕ ЭЛЕМЕНТОВ ОРГРАФА. АЛГОРИТМ ФАЛКЕРСОНА

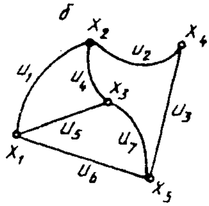
До сих пор мы изображали графы рисунками. При большом числе вершин и дуг (ребер) рисунок графа теряет наглядность. В таких случаях для задания графов и работы с ними используются матрицы. Особенно удобно использовать матричный эквивалент графа при его исследовании на ЭВМ.

а



$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\begin{array}{c} \begin{array}{ccccccc} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

б



$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\begin{array}{c} \begin{array}{ccccccc} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

Рис. 4.5

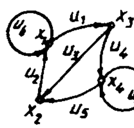
Рассмотрим граф G без петель, имеющий n вершин и m дуг. Графу G можно сопоставить *матрицу инциденций орграфа* G. Это прямоугольная матрица размерности $m \times n$, строки которой соответствуют вершинам, а столбцы — дугам графа. Элементы матрицы r_{ij} равны: -1, если дуга u_j исходит из i-й вершины; $+1$, если дуга заходит в i-ю вершину; 0, если дуга не инцидентна i-й вершине. В случае неориентированного графа элементами матрицы будут 1 и 0.

На рис. 4.5, *а* представлены орграф и его матрица инциденций, на рис. 4.5, *б* — неориентированный граф и его матрица инциденций.

Строки матрицы инциденций называют *векторами инциденций графа* G.

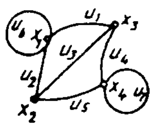
Как для орграфов, так и для неориентированных графов можно определить матрицу смежности вершин. *Матрица смежности вершин орграфа* G, содержащего n вершин, — это квадратная матрица $[p_{ij}]$ n-го порядка, строки и столбцы которой соответствуют вершинам графа G. Элементы p_{ij} матрицы равны числу дуг, идущих из i-й вершины в j-ю. Если орграф не содержит параллельных дуг, то матрица состоит только из 1 и 0.

а



$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline x_1 & 1 & 0 & 1 & 0 \\ x_2 & 1 & 0 & 0 & 0 \\ x_3 & 0 & 1 & 0 & 1 \\ x_4 & 0 & 1 & 0 & 1 \end{array}$$

б



$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline x_1 & 1 & 1 & 1 & 0 \\ x_2 & 1 & 0 & 1 & 1 \\ x_3 & 1 & 1 & 0 & 1 \\ x_4 & 0 & 1 & 1 & 1 \end{array}$$

Рис. 4.6

В случае неориентированного графа ему вместе с ребром (x_i, x_j) принадлежит и ребро (x_j, x_i), поэтому матрица будет симметрической. Справедливо и обратное утверждение: любой симметрической матрице с целыми неотрицательными элементами можно поставить в соответствие граф.

На рис. 4.6, *а* представлены орграф и его матрица смежности вершин, а на рис. 4.6, *б* — неориентированный граф со своей матрицей смежности.

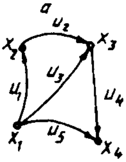
По матрице смежности вершин легко определить полустепени захода и исхода вершин, а значит, и степень вершины. Полустепень захода $P^+(x_i)$ вершины x_i равна сумме элементов i-го столбца; полустепень исхода $P^-(x_i)$ — сумме элементов i-й строки. Так, для графа, изображенного на рис. 4.6, *а*, по его матрице смежности находим, например, для вершины x_3: $P^+(x_3) = 1$, $P^-(x_3) = 2$, так что $P(x_3) = 3$.

В дальнейшем используется понятие непосредственного предшествования дуг. Будем говорить, что дуга u_i непосредственно предшествует дуге u_j, если конец дуги u_i является началом дуги u_j.

Граф G можно задать *матрицей смежности дуг* (*ребер*). Матрица смежности дуг орграфа — это квадратная матрица $[q_{ij}]$ m-го порядка (m — число дуг). Строки и столбцы матрицы соответствуют дугам (ребрам) графа. Элементы q_{ij} равны 1, если дуга u_i непосредственно предшествует дуге u_j, и 0 в остальных случаях. Матрицей смежности ребер неориентированного графа является матрица m-го порядка (m — число ребер) с элементами q_{ij}, равными 1, если ребра u_i и u_j смежны, и 0 в остальных случаях.

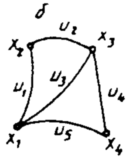
На рис. 4.7, *а* представлены орграф и его матрица смежности дуг, на рис. 4.7, *б* — неориентированный граф и его матрица смежности ребер.

а



$$\begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & u_3 & u_4 & u_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

б



$$\begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{array} \begin{array}{c} \begin{array}{ccccc} u_1 & u_2 & u_3 & u_4 & u_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

Рис. 4.7

Расчеты в задачах, связанных с графами, заметно упрощаются, если их элементы упорядоченны. Здесь потребуется понятие предшествования вершин. Говорят, что вершина x_i предшествует вершине x_j, если существует путь из x_i в x_j, тогда x_i называют предшествующей вершине x_j, а x_j — последующей за x_i. Под *упорядочением вершин* связного орграфа без контуров понимают такое разбиение его вершин на группы, при котором: 1) вершины первой группы не имеют предшествующих, а вершины последней — последующих; 2) вершины любой другой группы не имеют предшествующих в следующей группе; 3) вершины одной и той же группы дугами не соединяются. Можно показать, что описанное разбиение всегда возможно.

Аналогичным образом вводится понятие *упорядочения дуг*.

В результате упорядочения элементов получают граф, изоморфный данному. Упорядочение элементов выполняют графическим или матричным способом. Графический способ упорядочения вершин (*алгоритм Фалкерсона*) состоит из следующих шагов.

1. Находят вершины графа, в которые не входит ни одна дуга. Они образуют первую группу. Нумеруют вершины группы в натуральном порядке 1, 2, ... При этом присвоение номеров вершинам внутри группы может быть сделано не единственным образом, что не имеет значения.

2. Мысленно вычеркивают все пронумерованные вершины и дуги, из них выходящие. В получившемся графе найдется по крайней мере одна вершина, в которую не входит ни одна дуга. Этой вершине, входящей во вторую группу, присваивают очередной номер и т. д. Этот шаг повторяют до тех пор, пока все вершины не будут упорядоченны (пронумерованы).

Аналогичным образом упорядочивают дуги орграфа. Сначала находят дуги, не имеющие непосредственно предшествующих (они образуют I группу). После вычеркивания дуг I группы в оставшемся графе вновь выделяют дуги, не имеющие непосредственно предшествующих (они образуют II группу). И так до тех пор, пока все дуги не будут разбиты на группы. В заключение упорядоченным дугам присваивают новые обозначения с индексами 1, 2, ...

Пример 4.1. Упорядочить вершины данного графа и построить изоморфный граф.

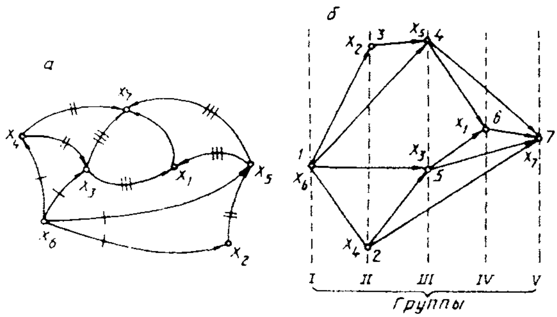


Рис. 4.8

Решение. Анализируя данный граф (рис. 4.8, *а*), замечаем, что в вершину x_6 не входит ни одна дуга. Следовательно, x_6 не имеет предшествующих, а потому относим ее к I группе. Больше подобных вершин на графе нет. Исключаем из рассмотрения вершину x_6 и дуги, из нее исходящие (на рис. 4.8, *а* эти дуги отмечены одной черточкой — первое вычеркивание). В оставшемся графе опять находим вершины, в которые не заходит ни одна дуга. Таковыми будут x_2 и x_4. Они образуют II группу. Выполняем второе вычеркивание и т. д. (см. рис. 4.8, *а*). На рис. 4.8, *б* для наглядности проведены вертикали, соответствующие группам разбиения, на которых последовательно отмечались точки: сначала x_6, затем x_2 и x_4 и т. д. В заключение эти точки соединили дугами так, как на данном графе, и получили изоморфный граф с упорядоченными вершинами. Остается перенумеровать его вершины в натуральном порядке.

Пример 4.2. Упорядочить дуги данного графа (рис. 4.9) и построить изоморфный граф.

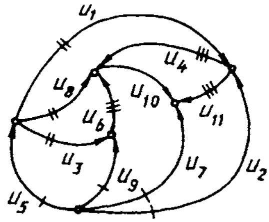


Рис. 4.9

Решение. К I группе относятся дуги u_2, u_5, u_7, u_9. После их вычеркивания непосредственно предшествующих не будут иметь дуги u_1, u_3 и u_8 (это дуги II группы). К III группе отойдут дуги u_4, u_6, u_{11} и к IV группе — дуга u_{10}.

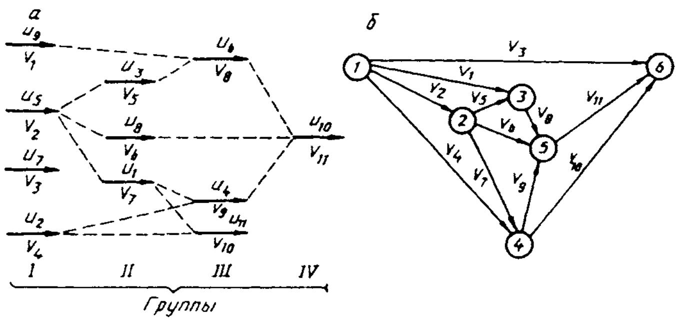


Рис. 4.10

На рис. 4.10, *а* изображены упорядоченные дуги данного графа. Они заново обозначены буквами $v_1, v_2, \ldots, v_{11}$. Штриховыми линиями показаны связи между дугами, существующие в данном графе. На рис. 4.10, *б* изображен граф, изоморфный данному, но с упорядоченными дугами.

4.3. ПОТОКИ НА СЕТЯХ. ПОСТАНОВКА ЗАДАЧИ О МАКСИМАЛЬНОМ ПОТОКЕ

Теория потоков возникла первоначально в связи с разработкой методов решения задач, связанных с рациональной перевозкой грузов. Схема доставки груза представлялась в виде графа, по ребрам которого проходит подлежащий максимизации поток этого груза. Позднее обнаружилось, что к задаче о максимальном потоке сводятся и другие важные оптимизационные практические задачи, такие, например, как задача отыскания минимального по стоимости плана выполнения комплекса работ при заданной его продолжительности; задача об определении максимального количества информации, которая может быть передана по разветвленной сети каналов связи из одного пункта в другой; задача об оптимальных назначениях; различные задачи организации снабжения; задачи автоматизации проектирования узлов вычислительных машин (печатный монтаж, трассировка плат); задачи, связанные с наиболее экономным строительством энергетических сетей, нефте- и газопроводов, железных и шоссейных дорог, ирригационных систем и много других прикладных задач.

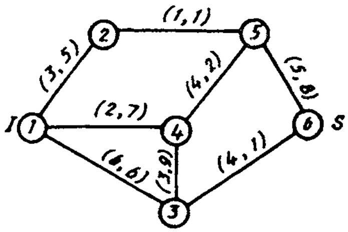


Рис. 4.11

Основным в теории потоков является понятие сети. *Сеть* — это взвешенный конечный граф без циклов и петель, ориентированный в одном общем направлении от вершины I, являющейся *входом (истоком) графа*, к вершине S, являющейся *выходом (стоком) графа* (рис. 4.11). Для наглядности будем представлять, что по ребрам (i, j) сети из истока I в сток S направляется некоторое вещество (груз, ресурс, информация и т. п.). В теории потоков предполагается, что если ребро (i, j) входит в сеть, то в сеть входит и ребро (j, i). Ребрам сети присваивается одна или несколько числовых характеристик.

Общее количество вершин сети будем обозначать n. На рис. 4.11 истоком I является вершина *1*, стоком S – вершина *6*.

Максимальное количество r_{ij} вещества, которое может пропустить за единицу времени ребро (i,j), называется его *пропускной способностью.* В общем случае $r_{ij} \neq r_{ji}$. Если вершины k и l на сети не соединены, то $r_{kl} = r_{lk} = 0$. На сети (рис. 4.11) пропускные способности ребер указаны в скобках. При этом первое число — это пропускная способность в направлении от вершины i к вершине j, второе — в противоположном направлении. Пропускные способности сети можно задать квадратной матрицей R n-го порядка. Поскольку $r_{ii} = 0$, на главной диагонали этой матрицы стоят нули. В табл. 4.1 приведена матрица R пропускных способностей сети, изображенной на рис. 4.11. Здесь $n = 6$, и поэтому матрица R — 6-го порядка.

Таблица 4.1

i \ j	*1*	*2*	*3*	*4*	*5*	*6*
1	0	3	6	2	0	0
2	5	0	0	0	1	0
3	6	0	0	3	0	4
4	7	0	9	0	4	0
5	0	1	0	2	0	5
6	0	0	1	0	8	0

Количество x_{ij} вещества, проходящего через ребро (i,j) в единицу времени, называется *потоком по ребру* (i,j).

Произвольно задать n^2 чисел x_{ij} $(i,j = \overline{1,n})$ нельзя. Они должны подчиняться определенным ограничениям, о которых речь пойдет дальше. А пока будем считать, что если поток из вершины i в вершину j равен x_{ij}, то поток из вершины j в вершину i будет равен $-x_{ij}$, т. е.

$$x_{ji} = -x_{ij}. \tag{4.1}$$

Принимается также, что $x_{ii} = 0$.

Если поток x_{ij} по ребру (i,j) меньше его пропускной способности, т. е. $x_{ij} < r_{ij}$, то ребро (i,j) называют *ненасыщенным*, если же $x_{ij} = r_{ij}$, — *насыщенным*.

Совокупность $X = \{x_{ij}\}$ потоков по всем ребрам (i, j) сети называют *потоком по сети* или просто *потоком.*

Из физического смысла грузопотока следует, что поток по каждому ребру (i, j) не может превышать его пропускную способность, т. е.

$$x_{ij} \leqslant r_{ij} \quad (i, j = \overline{1, n}). \tag{4.2}$$

Понятно также, что для любой вершины, кроме истока I и стока S, *количество вещества, поступающего в эту вершину, равно количеству вещества, вытекающего из нее.* С учетом соглашения (4.1) это требование можно выразить записью

$$\sum_{j=1}^{n} x_{ij} = 0 \quad (i \neq I, S). \tag{4.3}$$

Это ограничение называют *условием сохранения потока*: *в промежуточных вершинах потоки не создаются и не исчезают.* Отсюда следует, что *общее количество вещества, вытекающего из истока* I, *совпадает с общим количеством вещества, поступающего в сток* S, т. е.

$$f = \sum_{j} x_{Ij} = \sum_{i} x_{iS}, \tag{4.4}$$

где j — конечные вершины ребер, исходящих из I; i — начальные вершины ребер, входящих в S.

Линейную функцию f называют *мощностью потока* на сети.

Учитывая сказанное, задачу о максимальном потоке можно сформулировать следующим образом: *найти совокупность* $X^* = \{x^*_{ij}\}$ *потоков* x^*_{ij} *по всем ребрам* (i, j) *сети, которая удовлетворяет условиям (4.1) — (4.3) и максимизирует линейную функцию (4.4).* Это типичная задача линейного программирования.

Заметим, что числа x_{ij} образуют квадратную матрицу n-го порядка, на главной диагонали которой стоят нули $(x_{ii} = 0)$, а элементы, расположенные симметрично главной диагонали, равны по абсолютной величине и противоположны по знаку (см. соглашение (4.1)). Отсюда следует, что *задать поток* $X = \{x_{ij}\}$ *на сети — это значит задать* n^2 *чисел* x_{ij}, *удовлетворяющих условиям (4.1) — (4.3).*

Обратимся к рис. 4.11 и выясним, как практически можно организовать на сети какой-либо поток, удовлетворяющий условиям (4.1) — (4.3). С этой целью рассмотрим, например, полный путь *1 - 2 - 5 - 6*. (*Полным* называют путь из истока *I* в сток S.) Ребро (*2*, *5*), лежащее на этом пути, не позволяет пропустить более 1 ед. Следовательно, поток по указанному пути мощностью в 1 ед. будет допустимым, т. е. условия (4.1) — (4.3) выполняются для всех вершин и ребер этого пути. Проверим, например, условие (4.3) для вершины *2*:

$$x_{21} + x_{25} = (-x_{12}) + x_{25} = (-1) + 1 = 0.$$

По пути *1 - 4 - 5 - 6* можно пропустить 2 ед. (лимитирующим здесь является ребро (*1*, *4*)); по пути *1 - 3 - 6* пропустим 4 ед. (больше пропустить не позволяет ребро (*3*, *6*)). В результате потоки по ребрам будут равны: $x_{12} = 1$, $x_{13} = 4$, $x_{14} = 2$, $x_{25} = 1$, $x_{36} = 4$, $x_{45} = 2$, $x_{56} = 1 + 2 = 3$, а по остальным ребрам сети потоки равны нулю. Совокупность указанных потоков по ребрам и составит поток по сети. В соответствии с формулой (4.4) мощность сформированного потока $f = x_{12} + x_{13} + x_{14} = x_{36} + x_{56} = 7$ ед.

Чтобы ответить на вопрос: будет ли этот поток максимальным, необходимо его исследовать. Об этом речь пойдет дальше. А пока, учитывая соглашение (4.1), представим сформированный поток матрицей (табл. 4.2).

Таблица 4.2

i \ *j*	*1*	*2*	*3*	*4*	*5*	*6*
1	0	1	4	2	0	0
2	−1	0	0	0	1	0
3	−4	0	0	0	0	4
4	−2	0	0	0	2	0
5	0	−1	0	−2	0	3
6	0	0	−4	0	−3	0

4.4. РАЗРЕЗ НА СЕТИ. ТЕОРЕМА ФОРДА – ФАЛКЕРСОНА

Пусть дана некоторая сеть. Разобьем множество вершин сети на два непересекающихся подмножества A и B так, чтобы исток I попал в подмножество A, а сток S — в подмножество B. В этом случае говорят, что на сети произведен разрез, отделяющий исток I от стока S. В результате произведенного разбиения вершин появятся ребра (i, j), конечные точки которых окажутся в разных подмножествах. Совокупность ребер (i, j), начальные точки которых принадлежат подмножеству A, а конечные — подмножеству B, называют *разрезом сети* и обозначают A/B.

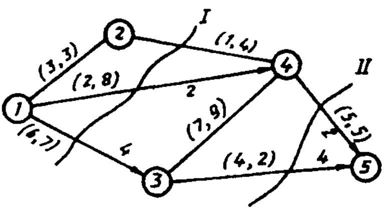


Рис. 4.12

На рис. 4.12 изображена сеть, на которой около каждого ребра указана его пропускная способность в обоих направлениях (в скобках) и величина потока по ребру. Стрелкой указано направление положительного потока. Кроме этого, на сети произведены два разреза: I и II. При разрезе I вершины сети оказались разбитыми на подмножества $A = \{1,\ 2\}$ и $B = \{3,\ 4,\ 5\}$, а ребрами, образующими разрез, стали $(1,\ 3)$, $(1,\ 4)$ и $(2,\ 4)$. При разрезе II $A = \{1,\ 2,\ 3,\ 4\}$; $B = \{5\}$, а образуют разрез ребра $(3,\ 5)$ и $(4,\ 5)$. Введем важные определения.

Величина

$$R(A/B) = \sum_{i \in A} \sum_{j \in B} r_{ij},$$

представляющая собой сумму пропускных способностей r_{ij} всех ребер разреза, называется *пропускной способностью разреза.*

Пусть на сети задан поток $X = \{x_{ij}\}$ и произведен разрез A/B. Величина

$$X(A/B) = \sum_{i \in A} \sum_{j \in B} x_{ij},$$

представляющая собой сумму потоков x_{ij} по всем ребрам разреза, называется *потоком через разрез*.

Для разреза I на рис. 4.12 $R(I) = r_{13} + r_{14} + r_{24} = 6 + 2 + 1 = 9$, $X(I) = x_{13} + x_{14} + x_{24} = 4 + 2 + 0 = 6$; для разреза II $R(II) = 9$, $X(II) = 6$.

Если на сети задан поток $X = \{x_{ij}\}$ и произведен разрез A/B, то хотя бы одно ребро любого полного пути, ведущего из истока I в сток S, будет обязательно принадлежать разрезу A/B. При этом величина потока по любому полному пути не превышает пропускную способность каждого его ребра, а потому величина X суммарного потока, устремленного из истока I в сток S, не может превысить пропускную способность любого разреза сети, т. е.

$$\sum_{i \in A} \sum_{j \in B} x_{ij} \leqslant \sum_{i \in A} \sum_{j \in B} r_{ij}. \tag{4.5}$$

Оказывается, если удастся построить на сети поток $X^* = \{x^*_{ij}\}$, величина которого равна пропускной способности некоторого разреза A/B, то этот поток будет максимальным, а разрез A/B обладает минимальной пропускной способностью.

В самом деле, пусть для потока $X^* = \{x^*_{ij}\}$ и разреза A^*/B^* выполняется равенство

$$\sum_{i \in A^*} \sum_{j \in B^*} x^*_{ij} = \sum_{i \in A^*} \sum_{j \in B^*} r_{ij}, \tag{4.6}$$

но максимальным является не X^*, а поток $\widetilde{X} = \{\widetilde{x}_{ij}\}$. В таком случае поток через разрез A^*/B^* потока $\widetilde{X}$ будет больше потока через этот же разрез потока X^*, т. е.

$$\sum_{i \in A^*} \sum_{j \in B^*} \widetilde{x}_{ij} > \sum_{i \in A^*} \sum_{j \in B^*} x^*_{ij}. \tag{4.7}$$

Объединив соотношения (4.6) и (4.7), получим

$$\sum_{i \in A^*} \sum_{j \in B^*} \widetilde{x}_{ij} > \sum_{i \in A^*} \sum_{j \in B^*} r_{ij},$$

что противоречит неравенству (4.5). Вторую часть высказанного выше утверждения примем без доказательства.

Проведенные рассуждения приводят к следующей теореме.

Теорема 4.1 **(Форда – Фалкерсона).** *На любой сети максимальная величина потока из истока I в сток S равна минимальной пропускной способности разреза, отделяющего I от S.*

Эта теорема имеет важное прикладное значение.

4.5. АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ О МАКСИМАЛЬНОМ ПОТОКЕ

Пусть на сети задан некоторый поток $X = \{x_{ij}\}$. Разобьем все вершины сети на два подмножества A и B следующим образом: к подмножеству A отнесем исток I и все вершины i, достижимые из I хотя бы по одному пути, состоящему из ненасыщенных ребер; к подмножеству B отнесем все остальные вершины, т. е. такие, которые нельзя достичь из I по ненасыщенным ребрам. При этом возможны следующие случаи: 1) сток $S \notin A$; 2) сток $S \in A$. Рассмотрим раздельно оба этих случая.

Случай 1. Если $S \notin A$, то $S \in B$, поэтому построенное разбиение является разрезом A/B. По условию разбиения для любой вершины $i \in A$ существует путь из I в i, состоящий из ненасыщенных ребер, а для любой вершины $j \in B$ такого пути нет. Отсюда следует, что любое ребро (i, j) разреза A/B (а у такого ребра $i \in A$, $j \in B$) будет насыщенным (иначе j принадлежало бы A), т. е. $x_{ij} = r_{ij}$. Просуммировав эти равенства по всем $i \in A$ и $j \in B$, получим

$$\sum_{i\in A}\sum_{j\in B} x_{ij} = \sum_{i\in A}\sum_{j\in B} r_{ij}. \tag{4.8}$$

В этом равенстве слева — величина X потока через разрез, справа — пропускная способность R разреза A/B. Из равенства (4.8) по теореме Форда — Фалкерсона следует, что поток $X = \{x_{ij}\}$ является максимальным.

Случай 2. Если $S \in A$, то существует путь из ненасыщенных ребер, ведущий из I в S. По ребрам этого пути можно пропустить дополнительный поток величиной $\Delta = \min(r_{ij} - x_{ij})$, где минимум берется по всем ребрам, входящим в этот путь. Потоки x_{ij} по всем остальным ребрам сети остаются прежними. В результате мощность суммарного потока возрастет на величину Δ. Это будет новый поток $X^1 = \{x_{ij}^1\}$.

Объединяя оба рассмотренных случая, можно предложить следующий алгоритм построения максимального потока.

1. Построить некоторый начальный поток $X^0 = \{x_{ij}^0\}$.

2. Организовать процедуру составления подмножества A вершин, достижимых из истока I по ненасыщенным ребрам. Если в этом процессе сток S не попадет в подмножество A, то построенный поток максимальный и задача решена. Если же S попал в A, то перейти к п. 3 алгоритма.

3. Выделить путь из I в S, состоящий из ненасыщенных ребер, и увеличить поток x_{ij} по каждому ребру этого пути на величину $\Delta = \min(r_{ij} - x_{ij})$, где минимум берется по ребрам (i, j) упомянутого пути. Тем самым будет построен новый поток $X^1 = \{x_{ij}^1\}$. После этого надо возвратиться к п. 2 алгоритма.

Заметим в заключение, что при выполнении п. 3 на каждом шаге по крайней мере одно из ненасыщенных ранее ребер становится насыщенным (именно то, которое соответствует Δ), а поскольку число ребер в сети конечно, через конечное число шагов максимальный поток будет построен.

Покажем, как реализуется изложенный алгоритм на практике.

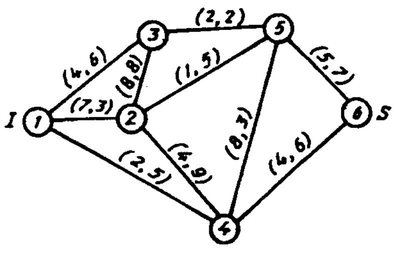


Рис. 4.13

На рис. 4.13 изображена сеть, у каждого ребра которой в скобках указаны два числа: первое означает пропускную способность ребра в направлении возрастания номеров, что совпадает с общим направлением грузопотока от I к S; второе — пропускная способность в противоположном направлении. В табл. 4.3 приведена матрица R пропускных способностей данной сети.

Таблица 4.3

i \ j	1	2	3	4	5	6
1	0	7	4	2	0	0
2	3	0	8	4	1	0
3	6	8	0	0	2	0
4	5	9	0	0	8	4
5	0	5	2	3	0	5
6	0	0	0	6	7	0

В соответствии с п. 1 алгоритма на сети необходимо сформировать какой-либо начальный поток. Примем в качестве такового поток X^0, в котором по пути *1 – 3 – 5 – 6* перемещаются 2 ед. (по ребру (*3*, *5*) больше пропустить нельзя!), по пути *1 – 2 – 5 – 6* 1 ед. (здесь лимитирующим является ребро (*2*, *5*)), по пути *1 – 4 – 6* 2 ед. (при этом ребро (*1*, *4*) становится насыщенным). Матрица потока X^0 приведена в табл. 4.4.

Таблица 4.4

i \ j	1	2	3	4	5	6
1	0	1	2	2	0	0
2	−1	0	0	0	1	0
3	−2	0	0	0	2	0
4	−2	0	0	0	0	2
5	0	−1	−2	0	0	3
6	0	0	0	−2	−3	0

Таблица 4.5

i \ j	1	2	3	4	5	6
1	0	6	2	0	0	0
2	4	0	8	4	0	0
3	8	8	0	0	0	0
4	7	9	0	0	8	2
5	0	6	4	3	0	2
6	0	0	0	8	10	0

В соответствии с формулой (4.4) мощность потока X^0

$$f = x_{12} + x_{13} + x_{14} = x_{46} + x_{56} = 1 + 2 + 2 = 2 + 3 = 5.$$

Приступая к выполнению п. 2 алгоритма, составим матрицу $R - X^0$ (табл. 4.5), элементы $r_{ij} - x_{ij}^0$ которой позволяют судить о насыщенности ребер сети. Насыщенным ребрам будут соответствовать нулевые элементы, а ненасыщенным — ненулевые. В нашем случае, например, ребро (*1*, *3*) ненасыщенное, поэтому элемент $r_{13} - x_{13}^0 = 4 - 2 = 2 \neq 0$, а вот ребро (*2*, *5*) насыщенное, поэтому $r_{25} - x_{25}^0 = 1 - 1 = 0$.

Зная матрицу $R - X^0$, можно сформировать подмножество A вершин, в которые можно попасть из истока I, двигаясь только по ненасыщенным путям (т. е. выполнить п. 2 алгоритма), а также выделить (если поток X^0 не максимален) эти пути и с их помощью увеличить мощность потока (т. е. выполнить п. 3 алгоритма).

Вершины подмножества A выделяют из всего множества вершин постепенно, начиная с истока I. С этой целью просматривают первую строку матрицы $R - X^0$ (строку I) и выписывают номера $i_1, i_2, \ldots, i_k$ вершин, соответствующих ненулевым элементам строки. Это и будут вершины, в которые можно попасть из истока I, перемещаясь по ненасыщенным ребрам. Будем записывать выявленные таким образом вершины в виде $I \| i_1, i_2, \ldots, i_k$ и называть подобную запись *списком вершины* I. Далее рассматривают каждую из вершин i_t полученного списка и составляют для нее аналогичным образом свой список. При этом вершины, встречавшиеся в прежних списках, повторно не выписываются.

Если в этом процессе сток S не встретится, то поток максимален и задача решена; если же при составлении очередного списка в нем появится сток S, то поток не максимален и мощность его можно увеличить.

Обратимся к рассматриваемому примеру. В данном случае $I \equiv 1$, $S \equiv 6$. Построим подмножество A, последовательно составляя списки вершин, начиная с вершины *1*. Судя по первой строке матрицы $R - X^0$ (см. табл. 4.5), в список вершины *1* войдут вершины *2* и *3*, так как элементы второго и третьего столбцов этой строки отличны от нуля. Итак, запишем: $1 \| 2, 3$. Теперь переходим к составлению списка вершины *2* как вершины, вошедшей в список вершины *1*. Во второй строке матрицы три элемента отличны от нуля: 4, 8, 4. Но 4 и 8 соответствуют вершинам *1* и *3*, которые уже значатся в подмножестве A, поэтому повторно их в списки не включаем. Элементу 4 четвертого столбца соответствует вершина *4*, которая встречается впервые, а потому включаем ее в список вершины *2*: $2\|4$. Следующая вершина, для которой надлежит составить список, — это вершина *3*. Однако в третьей строке матрицы $R - X^0$ ненулевым элементам 8 и 8 соответствуют вершины *1* и *2*, которые уже встречались в списках. Следовательно, список вершины *3* будет пустым: $3\|.$. Далее аналогичным образом составляется список вершины *4*: $4\|5, 6$. В результате получили следующий набор списков:

$$\boxed{1\|2,3}, \quad \boxed{2\|4}, \quad \boxed{3\|.}, \quad \boxed{4\|5,6}. \tag{4.9}$$

Анализируя списки (4.9), замечаем, что сток *6* попал в подмножество A, поскольку оказался в списке одной из вершин подмножества A (в данном случае вершины *4*). Значит, поток X^0 не максимален и существует путь из истока I в сток S (у нас из *1* в *6*), состоящий из ненасыщенных ребер. Таким образом, необходимо переходить к выполнению п. 3 алгоритма — выделению ненасыщенного пути из *1* к *6* и преобразованию с его помощью имеющегося на сети потока в новый поток большей мощности.

Продолжим общие рассуждения. Построение ненасыщенного пути из I в S начинают с последнего ребра этого пути. Понятно, что им будет ребро (i_{n-1}, S), где i_{n-1} — вершина, в список которой попал сток S. Далее выписывают ребро (i_{n-2}, i_{n-1}), где i_{n-2} — вершина, в список которой попала вершина i_{n-1}, и т. д., пока на очередном шаге не встретится ребро (I, i_1), которым и начинается искомый ненасыщенный путь. Итак, ненасыщенный путь из I в S состоит из ребер (I, i_1), $(i_1, i_2), \ldots, (i_{n-2}, i_{n-1})$, (i_{n-1}, S).

Вернемся к нашему примеру. Просматривая списки (4.9) от конца к началу, замечаем, что ребром (i_{n-1}, S) в нашем случае является (*4*, *6*), ребром (i_{n-2}, i_{n-1}) — ребро (*2*, *4*), ребром (I, i_1) — ребро (*1*, *2*). Таким образом, путь из истока *1* в сток *6* по ненасыщенным ребрам пройдет через вершины *1*, *2*, *4* и *6*.

После выделения ненасыщенного пути из истока I в сток S остается с помощью матрицы $R - X^0$ определить величину $\Delta = \min(r_{ij} - x_{ij})$, на которую нужно увеличить поток по каждому ребру (i, j) выделенного пути, чтобы получить новый поток X^1 мощности, большей на Δ единиц.

В нашем примере, как видно из табл. 4.5, по ребру (*1*, *2*) дополнительно можно пропустить 6 ед., по ребру (*2*, *4*) — 4 ед., а по ребру (*4*, *6*) — только 2 ед., поэтому увеличить поток по всему пути *1* – *2* – *4* – *6*, составленному из указанных ребер, можно лишь на 2 ед., так что

$$\Delta = \min_{1-2-4-6}(6, 4, 2) = 2.$$

Для построения матрицы нового потока X^1 к соответствующим элементам x_{ij}^0 матрицы X^0 прибавляют найденное значение Δ, после чего возвращаются к п. 2 алгоритма, и так до получения максимального потока.

Таблица 4.6

i \ j	*1*	*2*	*3*	*4*	*5*	*6*
1	0	3	2	2	0	0
2	−3	0	0	2	1	0
3	−2	0	0	0	2	0
4	−2	−2	0	0	0	4
5	0	−1	−2	0	0	3
6	0	0	0	−4	−3	0

Таблица 4.7

i \ j	*1*	*2*	*3*	*4*	*5*	*6*
1	0	4	2	0	0	0
2	6	0	8	2	0	0
3	8	8	0	0	0	0
4	7	11	0	0	8	0
5	0	6	4	3	0	2
6	0	0	0	10	10	0

В рассматриваемом примере на величину $\Delta = 2$ возрастут потоки x_{ij}^0 по ребрам (*1*, *2*), (*2*, *4*) и (*4*, *6*), составляющим ненасыщенный путь. По остальным ребрам сети величины потоков не изменятся. Итак, прибавляя 2 ед. к элементам $x_{12}^0 = 1$, $x_{24}^0 = 0$ и $x_{46}^0 = 2$ табл. 4.4 и учитывая соглашение (4.1), приходим к матрице нового потока X^1 (табл. 4.6), мощность которого равна 7 ед. Этот поток вновь надо исследовать на оптимальность, т. е. вернуться к п. 2 алгоритма. С этой целью, как и при исследовании потока X^0, составляем матрицу $R - X^1$ (табл. 4.7), а по ней — списки вершин, достижимых из истока *1* по ненасыщенным путям:

$$\boxed{1\|2,3}, \quad \boxed{2\|4}, \quad \boxed{3\|.}, \quad \boxed{4\|5}, \quad \boxed{5\|6}.$$

Из этих списков видно, что сток *6* оказался в подмножестве A, а путь, ведущий в него, состоит из ненасыщенных ребер (*1*, *2*), (*2*, *4*), (*4*, *5*), (*5*, *6*). Новый поток X^2 (табл. 4.8) получается преобразованием потока X^1 (см. табл. 4.6), если увеличить на

$$\Delta = \min_{1-2-4-5-6}(4,2,8,2) = 2$$

потоки по указанным ребрам найденного ненасыщенного пути. Мощность нового потока X^2 составляет 9 ед. Для исследования этого потока составлена матрица $R - X^2$ (табл. 4.9), а по ней — списки

$$\boxed{1 \| 2, 3}, \quad \boxed{2 \| .}, \quad \boxed{3 \| .}, \qquad (4.10)$$

Таблица 4.8

i \ j	1	2	3	4	5	6
1	0	5	2	2	0	0
2	−5	0	0	4	1	0
3	−2	0	0	0	2	0
4	−2	−4	0	0	2	4
5	0	−1	−2	−2	0	5
6	0	0	0	−4	−5	0

Таблица 4.9

i \ j	1	2	3	4	5	6
1	0	2	2	0	0	0
2	8	0	8	0	0	0
3	8	8	0	0	0	0
4	7	13	0	0	6	0
5	0	6	4	5	0	0
6	0	0	0	10	12	0

из которых видно, что сток *6* не попал в подмножество A вершин, достижимых из истока *1* по ненасыщенным путям. Значит, поток X^2 максимален. Остается нанести его на сеть с указанием направления потоков по отдельным ребрам (рис. 4.14).

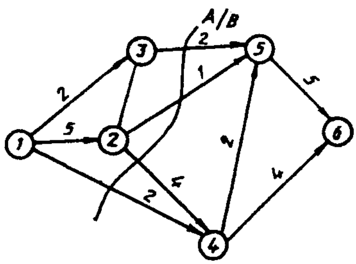


Рис. 4.14

Можно было и не строить матрицу $R - X^2$, если бы своевременно заметить, что потоки по ребрам (*4*, *6*) и (*5*, *6*) равны их пропускным способностям, т. е. эти ребра насыщенные (см. табл. 4.8 и 4.9).

Используя списки (4.10), выделим подмножества A и B, на которые оказалось разбитым множество всех вершин: $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$. А теперь можно выписать ребра, образующие разрез A/B минимальной пропускной способности: (*1*, *4*), (*2*, *4*), (*2*, *5*), (*3*, *5*).

4.6. ПРИЛОЖЕНИЯ ЗАДАЧИ О МАКСИМАЛЬНОМ ПОТОКЕ

Транспортная задача по критерию времени. Проблемы, возникающие при рациональной организации перевозок различных грузов, имеют исключительное народнохозяйственное значение. Математические задачи, к которым можно свести указанные проблемы, называют *транспортными*. Изучение методов решения транспортных задач важно еще и потому, что большое количество других прикладных задач можно описать математической моделью, сходной с моделью задачи о перевозках, а следовательно, и решить по аналогичным алгоритмам. Качество плана транспортной задачи можно оценивать по различным критериям. Здесь мы ограничимся лишь критерием времени. Поставим задачу.

Пусть известны запасы груза a_i $(i = \overline{1, m})$ у поставщиков A_i, спрос b_j $(j = \overline{1, n})$ потребителей B_j и время t_{ij} доставки груза (независимо от объема поставки) по маршруту $A_i - B_j$. Составить реализуемый за минимальное время план перевозок, при котором спрос удовлетворяется полностью.

Если суммарный запас груза совпадает с суммарным спросом, т. е. $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$, то задачу называют *закрытой*, в противном случае — *открытой*.

Составим математическую модель задачи. Обозначим через x_{ij} количество груза, планируемое к перевозке из i-го пункта поставки в j-й пункт потребления, через t — время наиболее продолжительной перевозки. Оптимальным будем считать план $(x_{11}, x_{12}, \ldots, x_{mn})$, самая продолжительная перевозка которого минимизируется. Модель закрытой задачи имеет вид:

$$\min t = \max_{x_{ij}>0} t_{ij}; \tag{4.11}$$

$$\sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1,m}), \tag{4.12}$$

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1,n}), \tag{4.13}$$

$$x_{ij} \geqslant 0 \quad (i = \overline{1,m};\ j = \overline{1,n}). \tag{4.14}$$

Как видно, целевая функция (4.11) является нелинейной. Уравнения (4.12) — ограничения по запасам — выражают требование, чтобы сумма всех поставок, идущих из i-го пункта, равнялась запасу a_i груза в нем; уравнения (4.13) — ограничения по потребностям — чтобы сумма всех поставок, направляемых в j-й пункт, равнялась его спросу b_j. Условие (4.14) означает, что обратные перевозки (возврат груза) не предполагаются.

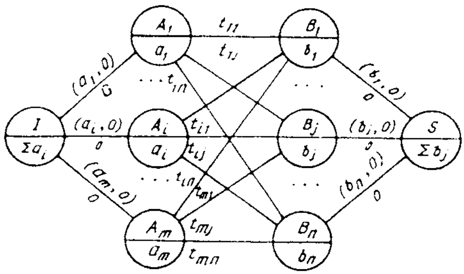


Рис. 4.15

Решение задачи (4.11) — (4.14) сведением ее к задаче о максимальном потоке является одним из известных методов. Суть метода в следующем. Строится сеть с $m+n+2$ вершинами, m из которых соответствуют поставщикам A_i, а n — потребителям B_j; две оставшиеся соответствуют истоку I и стоку S (рис. 4.15). Пропускные способности ребер полагают равными:

$$r_{IA_i} = a_i,\ r_{A_iI} = 0,\ r_{B_jS} = b_j,\ r_{SB_j} = 0,\ r_{A_iB_j} = r_{B_jA_i} = \infty.$$

У ребер (A_i, B_j) проставляются времена t_{ij} доставки груза. Время доставки по ребрам (I, A_i) и (B_j, S) считают равным нулю, т.е. $t_{IA_i} = t_{B_jS} = 0$. После построения сети по рассмотренной в § 4.5 методике отыскивается поток заданной мощности

$$f_{\max} = \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j,$$

при котором $\max t_{ij}$ достигает минимальной величины. В процессе этого поиска при наличии альтернативы исключаются из рассмотрения маршруты с более продолжительными поставками. Решение заканчивается, когда замена более продолжительных маршрутов менее продолжительными становится невозможной.

Задача об оптимальном назначении. Мы ограничимся рассмотрением упрощенного варианта задачи, сохраняющего, однако, все основные особенности общей задачи.

Пусть некоторая комплексная работа P связана с производством совокупности m более мелких работ $P_1, P_2, \ldots, P_m$, которые могут выполняться независимо одна от другой. В распоряжении планирующего органа находится n организаций-исполнителей $И_1$, $И_2, \ldots, И_n$, каждая из которых может выполнять только некоторые определенные работы. При этом каждый исполнитель одновременно может выполнять только какую-либо одну работу и каждая работа одновременно может выполняться только одним исполнителем. Задача состоит в распределении работ между исполнителями таким образом, чтобы одновременно выполнялось возможно большее их число.

Составим математическую модель задачи в предположении, что задана матрица $[a_{ij}]_{m \times n}$, элементы которой характеризуют возможности исполнителей, а именно: $a_{ij} = 1$, если i-я работа может выполняться j-м исполнителем, и $a_{ij} = 0$ в противном случае.

Обозначим через x_{ij} $(i = \overline{1, m};\ j = \overline{1, n})$ переменные, характеризующие распределение работ между исполнителями, и согласимся приписывать переменной x_{ij} значение, равное 1, если i-я работа поручена j-му исполнителю, и значение, равное 0, в противном случае. Итак,

$$x_{ij} = 1 \quad \text{или} \quad 0 \qquad (i = \overline{1,m};\ j = \overline{1,n}). \tag{4.15}$$

Поскольку каждому исполнителю можно поручить не больше одной работы, должно выполняться условие

$$\sum_{i=1}^{m} x_{ij} \leqslant 1 \quad (j = \overline{1,n}). \tag{4.16}$$

По условию задачи каждую работу можно поручить только такому исполнителю, который способен ее выполнить. Это можно выразить записью

$$x_{ij} \leqslant a_{ij} \quad (i = \overline{1,m};\ j = \overline{1,n}). \tag{4.17}$$

Поскольку каждая работа поручается не более чем одному исполнителю, то должно выполняться условие

$$\sum_{j=1}^{n} x_{ij} \leqslant 1 \quad (i = \overline{1,m}). \tag{4.18}$$

Очевидно, что общее число m работ, одновременно выполняемых всеми n исполнителями, можно представить в виде

$$f = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}. \tag{4.19}$$

Из математической модели (4.15) — (4.19) видно, что задача об оптимальных назначениях является задачей целочисленного линейного программирования, а потому может быть решена известными аналитическими методами.

Здесь мы рассмотрим решение задачи сведением ее к задаче о максимальном потоке. С этой целью строится сеть с $m + n + 2$ вершинами, m из которых соответствуют работам P_i, а n — исполнителям $И_j$, одна — истоку I и одна — стоку S (рис. 4.16). Исток I соединяют с вершинами P_i и считают пропускные способности ребер (I, P_i) равными 1, а ребер (P_i, I) равными 0. Вершины $И_j$ соединяют со стоком S и считают пропускные способности получившихся ребер $(И_j, S)$ равными 1, а ребер $(S, И_j)$ равными 0. Вершину P_i соединяют ребром с вершиной $И_j$ тогда и только тогда, когда $a_{ij} = 1$, т. е.

когда работа P_i может быть выполнена исполнителем $И_j$. При этом пропускную способность такого ребра $(P_i, И_j)$ считают равной 1, а ребра $(И_j, P_i)$ — равной 0.

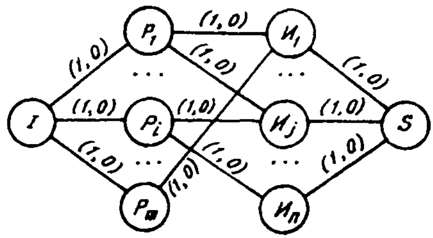


Рис. 4.16

Каждому распределению работ можно поставить в соответствие поток на сети. В самом деле, если работа P_i поручается исполнителю $И_j$, то по цепочке ребер (I, P_i), $(P_i, И_j)$, $(И_j, S)$ пропускается поток единичной мощности. Можно показать и обратное: любому потоку с целочисленными компонентами соответствует некоторое распределение работ. При таком соответствии число распределенных работ равно мощности суммарного потока из истока I в сток S. Поэтому, чтобы решить задачу об оптимальных назначениях, достаточно найти в соответствующей сети максимальный поток с целочисленными потоками по ребрам сети.

4.7. ЭЛЕМЕНТЫ СЕТЕВОГО ПЛАНИРОВАНИЯ

При планировании и оперативном управлении сложными комплексами работ, объединенных общностью цели, с успехом используются их графические модели — *сетевые графики (сети)*. С математической точки зрения сетевой график — это связный орграф без петель и контуров. В настоящее время разработаны специальные математические методы *сетевого планирования и управления* (СПУ). Основными понятиями СПУ являются работа и событие. Под *работой* понимаются любые действия, трудовые процессы, сопровождающиеся затратами ресурсов или времени и приводящие к определенным результатам. Под *событием* понимают результат завершения одной или нескольких работ. Событие является предпосылкой для выполнения работ, следующих за ним. Поэтому любая работа на сети может быть определена двумя событиями, между которыми она находится. Событием же может заканчиваться или начинаться сразу несколько работ. Работы на сети изображают произвольной длины направленными отрезками прямых (стрелками), а события — обычно кружками, в которых указывают порядковый номер или шифр события. У каждой стрелки проставляется время выполнения работы, а иногда и другие числовые характеристики (расход ресурса, количество исполнителей и т. д.). Сетевые графики выполняются с соблюдением определенных правил. В частности, он должен иметь только одно исходное событие (исток сети I) — начало работ комплекса — и только одно завершающее событие (сток сети S) — окончание всех работ комплекса. Но, прежде чем строить сеть, надо составить подробный список работ комплекса, в отношении каждой работы выяснить ее технологические связи с другими работами, место работы в комплексе, конечные результаты (события) каждой работы. После того как описанный подготовительный этап будет закончен, приступают к построению сети.

Пример 4.3. По данным табл. 4.10 построить сеть.

Таблица 4.10

Обозначение работы	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
Непосредственно предшествующие работы	—	—	—	a_1	a_1, a_2	a_1, a_2	a_3, a_5	a_4, a_6, a_7	a_3, a_5
Продолжительность работы	3	6	4	5	1	9	6	8	5

Решение. Работы a_1, a_2 и a_3 не имеют предшествующих, поэтому реализация комплекса начинается с этих работ. Они изображаются дугами, выходящими из одного кружка — события *1* (исток I сети) (рис. 4.17). Масштаб при этом не соблюдается и дуги a_1, a_2, a_3 располагаются достаточно произвольно. Работе a_4 предшествует работа a_1, поэтому дуга a_4 на сети изображена вслед за дугой a_1. Далее надо изобразить работы a_5 и a_6, выполняемые после работ a_1 и a_2. Во избежание путаницы на сетях не рекомендуется изображать параллельными дугами одновременно выполняемые работы. В подобных случаях условились вводить дополнительные события и фиктивные работы (нулевой продолжительности), которые изображаются штриховыми линиями. Их назначение —

показать, что одна работа не может быть выполнена ранее какого-либо события или работы. Учитывая сказанное, введем фиктивную работу, соединив конечное событие *2* работы a_1 с конечным событием *3* работы a_2, и после этого изобразим работы a_5 и a_6 дугами a_5 и a_6, выходящими из события *3*. Дуги a_3 и a_5 пришлось свести в одно событие *4*, поскольку работы a_7 и a_9 могут начаться лишь после завершения работ a_3 и a_5.

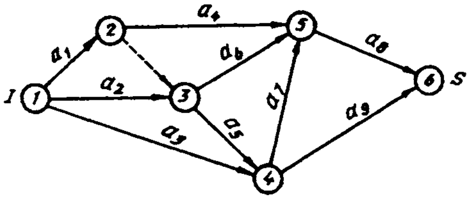


Рис. 4.17

Аналогичным образом поступим и с дугами a_4, a_6 и a_7, направив их в общее событие *5*, учитывая, что работа a_8 может быть начата только после выполнения работ a_4, a_6 и a_7. И, наконец, дуги a_8 и a_9 моделируют заключительные работы a_8 и a_9 комплекса, а потому сведем их в одно завершающее комплекс событие *6* (сток S сети). Правильность нумерации событий (вершин графа) следует проверить, воспользовавшись алгоритмом Фалкерсона (см. § 4.2).

Допустим, что сеть некоторого комплекса построена и известна продолжительность каждой работы (например, в днях) (рис. 4.18). Возникает вопрос: за какое минимальное время можно выполнить все работы комплекса? Чтобы ответить на

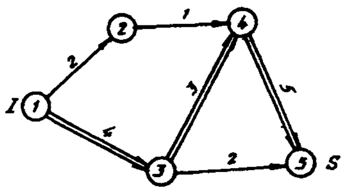


Рис. 4.18

этот вопрос, рассмотрим все полные пути L_i на сети. В данном случае таких путей три: L_1: *1 – 2 – 4 – 5*; L_2: *1 – 3 – 4 – 5*; L_3: *1 – 3 – 5*. Находим продолжительности $t(L_i)$ полных путей: $t(L_1) = 8$, $t(L_2) = 12$, $t(L_3) = 6$. Наиболее продолжительным оказался полный путь L_2. Его называют *критическим*. Он и определяет минимальное время выполнения всех работ данного комплекса. Это минимальное время называют *критическим сроком* и обозначают $t_{\text{кр}}$. В нашем случае $t_{\text{кр}} = 12$.

Работы и события, лежащие на критическом пути, называются *критическими*, остальные работы и события сети — *некритическими*. Если выполнение какой-либо критической работы будет задержано, это вызовет задержку выполнения всего комплекса на тот же срок. Чтобы ускорить выполнение комплекса, необходимо сократить сроки выполнения критических работ. Некритические работы допускают некоторое запаздывание их выполнения без нарушения критического срока. Чтобы определить время, на которое можно задержать выполнение некритических работ, вводят понятия *резервов времени событий* и *работ*, которые в свою очередь выражаются через ранние и поздние сроки свершения событий.

Как отмечалось, в событие может входить и выходить из него несколько работ. Под *свершением события* будем понимать момент, к которому заканчиваются все входящие в него работы и может быть начата любая выходящая работа. Событие может иметь некоторый интервал свободы свершения. Поясним эту мысль. Событие *2* на рис. 4.18 может свершиться через 2 дня (по окончании работы (*1*, *2*)). Но оно может наступить и позже, если добавить время на выполнение работы (*1*, *2*). А это сделать можно, поскольку на полном пути L_1, на котором лежит эта работа, есть резерв времени $R(L_1) = t_{\text{кр}} - t(L_1) = 12 - 8 = 4$. Так что работу (*1*, *2*) можно выполнять даже за $2 + 4 = 6$ дней, и это не повлияет на критический срок. Но тогда событие *2* наступит через 6 дней. Именно поэтому для событий различают ранний и поздний сроки свершения.

Ранним сроком $t_p(j)$ свершения события j назовем самый ранний момент времени, к которому завершаются все работы, предшествующие этому событию. Чтобы получить общую расчетную формулу, обратимся к рис. 4.18. Согласимся, что $t_{\text{р}}(1) = 0$, а $t_{\text{р}}(5) = t_{\text{кр}} = 12$. И вообще, $t_{\text{р}}(I) = 0$, $t_{\text{р}}(S) = t_{\text{кр}}$.

Событие *2* произойдет по завершении работы (*1*, *2*), т. е. через два дня. Значит, $t_{\text{р}}(2) = 2$. Запишем это иначе: $t_{\text{р}}(2) = 0 + 2 = t_{\text{р}}(1) + t(1, 2)$, где $t(1, 2)$ — время выполнения работы (*1*, *2*). Аналогично $t_{\text{р}}(3) = 0 + 4 = t_{\text{р}}(1) + t(1, 3)$. Немного сложнее с событием *4*, поскольку им завершается несколько работ (две). По работе (*2*, *4*) аналогично предыдущему можно записать: $t_р(2) + t(2,4) = 2 + 1 = 3$, а по работе (*3*, *4*) — $t_р(3) + t(3,4) = 4 + 3 = 7$. Поскольку к моменту $t_р(4)$ должны закончиться все предшествующие работы, понятно, что $t_р(4)$ следует выбрать из условия $t_р(4) = \max(t_р(2) + t(2,4), t_р(3)+$ $+t(3,4))$. И мы приходим к формуле

$$t_р(j) = \max_{(i,j)\in U_j^+}(t_р(i) + t(i,j)), \tag{4.20}$$

где U_j^+ — множество работ, входящих в j-е событие; $t_р(i)$ — ранний срок свершения начального события работы (i,j); $t(i,j)$ — продолжительность работы (i,j).

Итак, если событием j заканчивается одна работа (i,j), то $t_р(j)$ равен сумме $t_р(i)$ — раннего срока свершения ее начального события — и $t(i,j)$ — продолжительности этой работы; если же событием j заканчивается несколько работ, то по каждой работе находится свой ранний срок, а искомым $t_р(j)$ будет максимальный из них.

Поздним сроком $t_п(i)$ *свершения события* i назовем самый поздний момент времени, после которого остается ровно столько времени, сколько необходимо для завершения всех работ, следующих за этим событием.

Можно согласиться с тем, что $t_п(I) = 0$, $t_п(S) = t_р(S) =$ $= t_{кр}$.

В нашем примере $t_{кр} = 12$, поэтому $t_п(5) = 12$. Чтобы не нарушился критический срок, событие *4* должно произойти в крайнем случае на 5 дней раньше события *5* (5 дней необходимо для выполнения работы (*4*, *5*)), поэтому $t_п(4) =$ $= 12 - 5 = t_п(5) - t(4,5) = 7$. Аналогично для события *3*: по работе (*3*, *5*) можно записать: $12 - 2 = t_п(5) - t(3,5) = 10$, а по работе (*3*, *4*) $7 - 3 = t_п(4) - t(3,4) = 4$. Поскольку $t_п(3)$ — это предельный момент времени, после которого должны быть выполнены все последующие работы, то за $t_п(3)$ придется принять $\min(t_п(4) - t(3,4); t_п(5) - t(3,5)) = \min(4,10) = 4$. Если взять 10 и учесть, что для выполнения работ (*3*, *4*) и (*4*, *5*) потребуется еще $3 + 5 = 8$ дней, то общий срок реализации комплекса возрастет до 18 дней (вместо $t_{кр} = 12$). Обобщая результат, приходим к расчетной формуле

$$t_п(i) = \min_{(i,j)\in U_i^-}(t_п(j) - t(i,j)), \tag{4.21}$$

где U_i^- — множество работ, выходящих из i-го события; $t_\text{п}(j)$ — поздний срок свершения конечного события работы (i, j).

Итак, если событием i начинается одна работа (i, j), то $t_\text{п}(i)$ равен разности $t_\text{п}(j)$ — позднего срока свершения ее конечного события — и $t(i, j)$ — продолжительности этой работы; если же событием i начинается несколько работ, то по каждой работе находят свой поздний срок, а искомым $t_\text{п}(i)$ будет минимальный из них.

Разность между поздним и ранним сроками свершения события i составляет резерв $R(i)$ времени этого события:

$$R(i) = t_\text{п}(i) - t_\text{р}(i).$$

Резерв $R(i)$ показывает, на какой предельно допустимый срок может задержаться свершение события i без изменения срока наступлении завершающего события S. Понятно, что у критических событий ранние и поздние сроки свершения совпадают, так что резерв времени у них равен нулю.

Зная сроки свершения событий, для работы (i, j) можно найти: ранний срок $t_\text{р.н}(i, j)$ начала работы, ранний срок $t_\text{р.о}(i, j)$ окончания работы, поздний срок $t_\text{п.о}(i, j)$ окончания работы, поздний срок $t_\text{п.н}(i, j)$ начала работы:

$$t_\text{р.н}(i, j) = t_\text{р}(i), \quad t_\text{р.о}(i, j) = t_\text{р}(i) + t(i, j),$$

$$t_\text{п.о}(i, j) = t_\text{п}(j), \quad t_\text{п.н}(i, j) = t_\text{п}(j) - t(i, j).$$

Для работ отметим два основных вида резервов времени: полный резерв $R_\text{п}(i, j)$ и свободный резерв $R_\text{с}(i, j)$.

Полный резерв времени работы — это максимальное количество времени, на которое можно задержать начало работы или увеличить ее продолжительность, не нарушая критический срок.

Можно показать, что

$$R_\text{п}(i, j) = t_\text{п}(j) - t_\text{р}(i) - t(i, j). \tag{4.22}$$

Формулу (4.22) проиллюстрируем рис. 4.19.

Отдельные работы, помимо полного резерва, имеют свободный резерв времени, составляющий часть полного резерва, остающуюся после исключения резерва времени $R(j)$ конечного события j данной работы. Можно показать, что

$$R_{\text{c}}(i,j) = t_{\text{p}}(j) - t_{\text{p}}(i) - t(i,j).$$

Свободный резерв времени работы — это запас времени, на который можно отсрочить начало работы или увеличить ее продолжительность при условии, что она начнется в свой ранний срок и при этом ранние сроки начала последующих работ не изменятся, а потому комплекс завершится в критический срок.

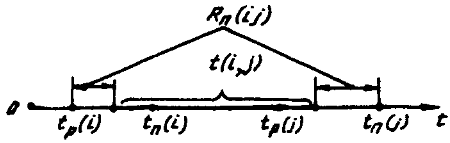


Рис. 4.19

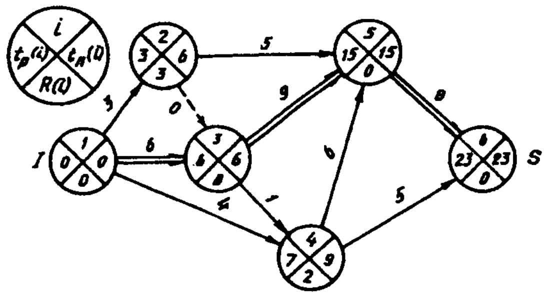


Рис. 4.20

Понятно, что о резервах времени можно говорить лишь применительно к некритическим работам, резервы критических работ равны нулю.

Существуют различные способы расчета временны́х параметров сети. При небольших комплексах расчеты производятся вручную, при значительных (свыше 500 событий) — на ЭВМ. Мы ограничимся способом расчета параметров непосредственно на сети и разберем его на примере сети, приведенной на рис. 4.17. С этой целью каждое событие изобразим кружком, разделенным диаметрами на четыре сектора (рис. 4.20). В верхнем секторе запишем номер i события, в левом по мере вычислений будем записывать ранний срок $t_р(i)$ свершения события i, в правом — поздний срок $t_п(i)$, в нижнем — резерв $R(i)$ времени события. Продолжительности работ возьмем из табл. 4.10.

Расчеты выполняют в четыре этапа, а именно вычисляют: 1) $t_р(i)$; 2) $t_п(i)$; 3) $R(i)$; 4) критический путь.

I этап. При вычислении $t_р(i)$ перемещаются по сети от события I к событию S в порядке возрастания номеров. Поскольку $t_р(I) = 0$, в левый сектор кружка *1* записываем 0. Затем рассматриваем событие *2*, в которое входит работа (*1*, *2*). В соответствии с формулой (4.20) $t_р(2) = t_р(1) + t(1,2) = 0 + 3 = 3$; это число и записываем в левый сектор кружка *2*. При вычислении $t_р(3)$ учитываем, что в событие *3* входят две работы: (*1*, *3*) и (*2*, *3*). (Особо подчеркнем, что во всех расчетах фиктивные работы учитываются наряду с реальными!) Поэтому по формуле (4.20) получаем

$$t_р(3) = \max(t_р(1) + t(1,3),\ t_р(2) + t(2,3)) =$$

$$= \max(0 + 6,\ 3 + 0) = 6,$$

что и записываем в левый сектор кружка *3*. Аналогично вычисляем ранние сроки и остальных событий, в том числе $t_р(6) = 23$, т. е. критический срок. Итак, $t_{кр} = 23$.

II этап. При вычислении поздних сроков свершения событий $t_п(i)$ перемещаются по сети от события S к событию I в порядке убывания номеров. Поскольку $t_п(S) = t_р(S)$, в правый сектор кружка *6* записываем число $t_п(6) = 23$. Рассматриваем предшествующее событие *5*, из которого выходит только одна работа (*5*, *6*). Следовательно, по формуле (4.21) находим $t_п(5) = t_п(6) - t(5,6) = 23 - 8 = 15$. Этот результат и записываем в правый сектор кружка *5*. Из события *4* выходят две работы (*4*, *5*) и (*4*, *6*), поэтому

$$t_п(4) = \min(t_п(5) - t(4,5),\ t_п(6) - t(4,6)) =$$

$$= \min(15 - 6,\ 23 - 5) = 9,$$

что и записываем в правый сектор кружка *4*. Аналогично определяются поздние сроки свершения всех остальных событий сети. Заметим только, что результатом расчетов должно быть равенство $t_п(1) = t_р(1) = 0$.

III этап. Для определения резервов времени событий $R(i) = t_п(i) - t_р(i)$ достаточно из чисел, записанных в правых секторах кружков, вычесть числа, записанные в левых секторах, и заполнить нижние секторы.

IV этап. У критических событий резерв времени равен 0. В нашем примере критическими являются события *1*, *3*, *5* и *6*, они и определяют критические работы и критический путь *1 - 3 - 5 - 6*.

Все остальные временные параметры (сроки начала и окончания работ, резервы времени работ) выражаются через $t_р$ и $t_п$, а поэтому легко могут быть вычислены. Так, например,

$$t_{р.\,о}(2,5) = t_р(2) + t(2,5) = 3 + 5 = 8,$$

$$R_п(3,4) = t_п(4) - t_р(3) - t(3,4) = 9 - 6 - 1 = 2.$$

В рассмотренном примере критический путь на сети оказался единственным. Их может быть несколько. Заметим еще, что критический путь может включать и фиктивные работы.

При анализе и оптимизации комплекса работ наряду с сетевым графиком с успехом применяется *линейный график*. Построить его можно по данному сетевому графику или матрице, которой он задан. Линейный график позволяет решать целый ряд задач, связанных с комплексом работ.

Пример 4.4. Комплекс работ представлен сетевым графиком (рис. 4.21). Известны продолжительность t_{ij} (в днях) каждой работы и количество r_{ij} единиц ресурса, необходимое для выполнения работы в единицу времени, — интенсивность потребления ресурса (число в скобках). Построить линейный график и найти по нему критический срок, критические работы и резервы времени некритических работ. Построить шкалу потребления ресурса.

Решение. Каждая работа (i, j) на линейном графике изображается в привязке к оси времени Ot прямолинейным отрезком, длина которого в выбранном масштабе равна продолжительности t_{ij} ее выполнения (рис. 4.22). Поэтому время у отрезков не проставляется, но указывается интенсивность r_{ij} потребления ресурса. Работы изображаются в той же последовательности, что и на сети.

В нашем случае комплекс начинается работами (*1*, *2*) и (*1*, *3*), поэтому начала отрезков *1 - 2* и *1 - 3* расположим на вертикали $t = 0$ (в произвольных точках), а длины возьмем равными соответственно 3 и 8 ед. После работы (*1*, *2*) выполняются работы (*2*, *3*), (*2*, *4*) и (*2*, *5*), поэтому начала всех трех отрезков *2 - 3*, *2 - 4* и *2 - 5* следует взять на вертикали $t = 3$, а длины их будут равны соответственно 4, 4 и 5 ед. Работа (*3*, *5*) выполняется после завершения двух работ: (*1*, *3*) и (*2*, *3*), поэтому начало отрезка *3 - 5* придется расположить на вертикали $t = 8$

(а не $t = 7$). Начало последней работы (4, 5), следующей за работой (2, 4), находится на вертикали $t = 7$. График построен.

Найдем критический срок $t_{кр}$ и критические работы. В нашем случае последней является работа (4, 5), ее конечной точке 5 соответствует на оси времени Ot отметка $t = 14$, которая и определяет критический срок.

Ясно, что работа (4, 5), будучи заключительной работой комплекса, является критической. Непосредственно ей предшествует работа (2, 4), а этой работе — работа (1, 2); обе эти работы также являются критическими. Все остальные работы — некритические. Критические работы на рис. 4.22 выделены двойной чертой.

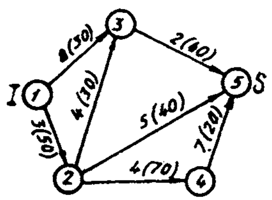


Рис. 4.21

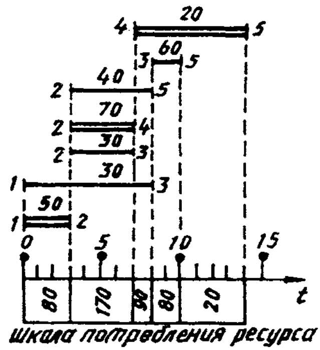


Рис. 4.22

По линейному графику нетрудно найти полные резервы времени $R_п(i, j)$ некритических работ. Так, работу (3, 5) в случае необходимости можно отсрочить или увеличить время ее выполнения на 4 дня (отрезок 3 – 5 на графике можно сдвинуть вправо на 4 ед. или "растянуть" его на 4 ед.), и это не вызовет нарушения $t_{кр} = 14$. Значит, $R_п(3, 5) = 4$. Аналогично $R_п(2, 5) = 6$. Что касается работы (2, 3), то через день после нее выполняется работа (3, 5), которую, как установлено, можно сдвинуть на 4 дня. В связи с этим при необходимости работу (2, 3) можно сдвинуть в общей сложности на 5 дней. Итак, $R_п(2, 3) = 5$. Аналогично определяется $R_п(1, 3) = 4$.

Определяя величину $R_п(2, 3)$, мы учли возможность сдвига работы (3, 5). Если же оперировать только с данной работой и совсем не затрагивать последующие работы, то найдется свободный резерв времени данной работы. В случае работы (2, 3) ее свободный резерв составляет всего 1 день, т. е. $R_с(2, 3) = 1$. Из рис. 4.22 непосредственно видно, что $R_с(1, 3) = 0$, $R_с(2, 5) = 6$, $R_с(3, 5) = 4$.

Чтобы проследить, как меняется интенсивность потребления ресурса в ходе работ, спроецируем на ось Ot начальные и конечные точки работ. Получим промежутки постоянства интенсивности: (0; 3); (3; 7); (7; 8);... Остается просуммировать в этих промежутках интенсивности r_{ij} работ, расположенных над ними: в промежутке (0; 3) суммарная интенсивность составляет $r_{12} + r_{13} = 50 + 30 = 80$ (ед.), в (3; 7) — $30+30+70+40 =$ $= 170$ (ед.) и т. д. Получили шкалу потребления ресурса.

Одной из наиболее распространенных оптимизационных задач сетевого планирования является задача о сокращении срока выполнения комплекса работ при ограниченных ресурсах. Она возникает в случаях, когда для реализации комплекса работ в плановый срок имеющихся ресурсов недостаточно. В такой обстановке приходится пересматривать сроки выполнения работ, переносить отдельные работы, сдвигая их во времени. При этом продолжительность выполнения комплекса, как правило, увеличивается, в связи с чем требуется произвести работы в новые сроки при имеющихся ресурсах в минимально возможное время. Здесь возникает вопрос: какие же работы целесообразнее отсрочить? Ответить на этот вопрос можно, полагаясь на опыт, здравый смысл, интуицию.

Пусть, например, за некоторый промежуток времени предполагается выполнить две работы, причем одна из них уже началась ранее. Выясняется, что одновременно вести обе работы невозможно из-за нехватки ресурсов. В подобной ситуации представляется естественным отсрочить (сдвинуть за пределы рассматриваемого промежутка) именно вторую работу, а ту, которая начата, оставить в первоначальном положении и продолжать.

Другой пример. В рассматриваемом промежутке должны начаться две работы, но вести их одновременно нет возможности. В этом случае целесообразнее оставить в первоначальном положении ту работу, у которой меньше полный резерв времени, а сдвинуть работу с бóльшим резервом. При равенстве резервов следует оставить работу с бóльшей интенсивностью, а работу с меньшей интенсивностью сдвинуть.

Мы говорили о двух работах. Если же предполагалось вести одновременно несколько работ, а это невозможно, то для выявления работ, подлежащих отсрочке, придется, руководствуясь приведенными соображениями, упорядочить эти работы, присвоив им соответствующие номера, и в порядке номеров проследить за тем, какая же работа вызывает превышение имеющегося в наличии ресурса; ее и сдвинуть.

Пример 4.5. Предположим, что в условиях примера 4.4 требуется установить время начала и окончания каждой работы так, чтобы завершить комплекс в возможно меньшее время при условии, что в любой момент в период выполнения работ расход ресурса не должен превышать $R = 100$ ед.

Решение. Из рис. 4.22 видно, что в промежутке (0; 3) суммарное потребление ресурса меньше имеющегося объема ($80 < 100$), поэтому работы (*1*, *2*) и (*1*, *3*) остаются в первоначальном положении и оптимизационный процесс начинается анализом промежутка (3; 7), где потребность в ресурсе превышает допустимый уровень ($170 > 100$).

Первый шаг. Над промежутком (3; 7) расположены четыре работы. Все их выполнять одновременно нет возможности. Для выявления работ, подлежащих отсрочке, прежде всего упорядочим их. Работа (*1*, *3*) начата раньше, поэтому она получает № 1. Для остальных работ сравниваем полные резервы времени: $R_{\text{п}}(2,3) = 5$, $R_{\text{п}}(2,4) = 0$, $R_{\text{п}}(2,5) = 6$. В порядке их возрастания присваиваем номера: работе (*2*, *4*) — № 2, работе (*2*, *3*) — № 3, работе (*2*, *5*) — № 4.

Для выявления работы, подлежащей сдвигу, выясним, какая из них вызывает превышение наличного запаса ресурса. С этой целью в порядке возрастания номеров суммируем интенсивности r_{ij} работ: $r_{13} + r_{24} = 30 + 70 = 100$. Поскольку имеющийся запас ресурса не превышен, работы (*1*, *3*) и (*2*, *4*) оставляем в прежнем положении, а суммирование продолжаем: $r_{13} + r_{24} + r_{23} = 30 + 70 + 30 = 130 > 100$. Превышение вызвано третьим слагаемым, поэтому соответствующую ему работу (*2*, *3*) надлежит сдвинуть вправо на длину рассматриваемого промежутка (3; 7), т. е. до момента $t = 7$.

Остается решить вопрос о четвертой работе (*2*, *5*). Возвращаясь к суммированию интенсивностей, отбрасываем слагаемое r_{23}, вызвавшее превышение запаса ресурса, а вместо него прибавляем интенсивность следующей по порядку номеров работы (*2*, *5*) r_{25}: $r_{13} + r_{24} + r_{25} = 30 + 70 + 40 = 140 > 100$. Аналогично предыдущему заключаем, что работа (*2*, *5*) также подлежит сдвигу до момента $t = 7$.

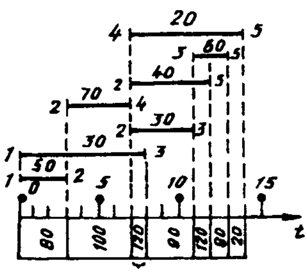


Рис. 4.23

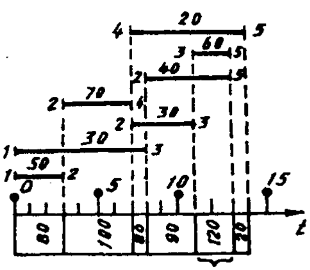


Рис. 4.24

Из рис. 4.22 видно, что сдвиг работы (*2*, *3*) на длину промежутка (3; 7), т. е. на 4 дня, вызывает смещение следующей за ней через 1 день работы (*3*, *5*) на 3 дня. Сдвиг работы (*2*, *5*) последствий не вызывает, так как за ней других работ нет.

Смещение работ повлечет за собой и преобразование шкалы потребления ресурса. На рис. 4.23 приведен преобразованный линейный график с соответствующей шкалой потребления ресурса. Первый шаг оптимизации закончен.

Второй шаг. Анализируя шкалу потребления ресурса (см. рис. 4.23), замечаем, что теперь исследованию подлежит промежуток (7; 8), так как до момента $t = 7$ запаса ресурса достаточно для выполнения работ. По сравнению с первым шагом здесь в рассуждениях ничего нового нет. Принимая во внимание, что работа (*1*, *3*) началась раньше, и учитывая полные резервы времени работ: $R_п(2,3) = 1$, $R_п(2,5) = 2$, $R_п(4,5) = 0$, присваиваем номера 1, 2, 3 и 4 соответственно работам (*1*, *3*), (*4*, *5*), (*2*, *3*) и (*2*, *5*). По результатам суммирования интенсивностей заключаем, что сдвигу до $t = 8$ подлежит работа (*2*, *5*). На рис. 4.24 приведен преобразованный линейный график вместе с изменившейся шкалой потребления ресурса. Тем самым завершен второй шаг оптимизации.

Третий шаг. Из рис. 4.24 видно, что теперь исследованию подлежит промежуток (11; 13). Над ним расположены три работы, две из которых — (*2*, *5*) и (*4*, *5*) — начаты ранее. Их и следует пронумеровать в первую очередь. Так как $R_п(2,5) = 1$, а $R_п(4,5) = 0$, то работе (*4*, *5*) присваиваем № 1, а работе (*2*, *5*) — № 2. Оставшейся работе (*3*, *5*), естественно, присваивается № 3. Суммированием интенсивностей устанавливаем, что работа (*3*, *5*) подлежит отсрочке до момента $t = 13$.

На рис. 4.25 приведен вновь преобразованный линейный график с соответствующей шкалой потребления ресурса. По шкале видно, что в любой момент реализации комплекса запаса ресурса достаточно для выполнения работ. В связи с этим процесс оптимизации заканчивается. Время, необходимое для производства всех работ, составляет 15 дней.

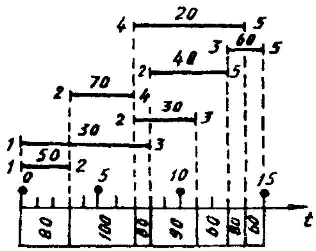


Рис. 4.25

По линейному графику (рис. 4.25) составлена табл. 4.11, в которой помещены новые сроки начала ($t_н$) и окончания ($t_о$) работ комплекса.

Таблица 4.11

Сроки	Работы						
	(*1*, *2*)	(*1*, *3*)	(*2*, *4*)	(*2*, *3*)	(*4*, *5*)	(*2*, *5*)	(*3*, *5*)
$t_н$	0	0	3	7	7	8	13
$t_о$	3	8	7	11	14	13	15

Рассмотренный при решении примера 4.5 эвристический метод оптимизации по времени комплекса работ, когда ресурсы ограничены, не обязательно точно минимизирует время выполнения комплекса работ, но обеспечивает достаточно хорошее приближение к нему.

Использованный метод кратко можно описать следующим образом. Анализируя шкалу потребления ресурса линейного графика комплекса работ, выделяют первый слева временной промежуток, в котором суммарная потребность в ресурсе для одновременного производства всех работ, расположенных над ним, превышает имеющийся запас R ресурса.

Затем определяют работы, подлежащие отсрочке (сдвигу). Для этого все работы упорядочивают по возрастанию полных резервов времени, а при их равенстве — по убыванию интенсивностей потребления ресурса. При этом первые номера отдают работам, начатым ранее анализируемого промежутка (если таковые имеются). Далее нумеруют работы, начинающиеся в анализируемом промежутке.

После упорядочения работ производят последовательное (по возрастанию номеров) суммирование интенсивностей r_{ij} потребления ресурса. Как только суммарная интенсивность превысит имеющийся запас R ресурса, слагаемое, вызвавшее превышение, отбрасывают, а соответствующую ему работу назначают к отсрочке на величину анализируемого промежутка, если работа начинается в этом промежутке, и до совмещения начала работы с моментом завершения анализируемого промежутка, если работа начинается левее его. После этого суммирование продолжают, пока вновь не обнаружится превышение R. И так до полного перебора всех работ над промежутком. Сроки выполнения работ, для которых суммарная интенсивность не превышает R, не меняются.

После завершения анализа производят преобразование линейного графика: сдвигают назначенные к отсрочке работы и работы, следующие за ними. Затем строят новую шкалу потребления ресурса. Этим завершается первый шаг оптимизационного процесса, в результате которого в рассмотренном временно́м промежутке потребление ресурса уже не превышает имеющегося запаса R.

Если на шкале потребления ресурса преобразованного линейного графика имеются промежутки, в которых суммарная потребность в ресурсе превышает R, то выполняют второй шаг оптимизационного процесса: выбирают самый левый из упомянутых промежутков и анализируют его аналогично предыдущему.

Описанную процедуру продолжают до тех пор, пока на шкале потребления ресурса не останется промежутков, в которых суммарное потребление превышает имеющийся запас R.

После завершения процесса оптимизации получают линейный график, по которому известным способом можно выделить критический путь. Он, как правило, отличается от критического пути исходного графика составляющими его работами. Изменится (увеличится) и продолжительность нового критического пути (критический срок). Это неизбежное следствие ограничения ресурса, используемого при производстве работ комплекса.

Так, на рис. 4.25 критический путь образуют работы (*1*, *3*), (*2*, *5*) и (*3*, *5*), а продолжительность критического пути составляет 15 дней (сравните с рис. 4.22).

Мы рассмотрели пример использования сетевых (линейных) графиков для решения оптимизационной задачи с достаточно простыми условиями. На практике нередко возникают гораздо более сложные задачи, когда требуется, например, определить работы, на производство которых необходимо выделить дополнительные ресурсы с целью форсирования этих работ, с тем чтобы критический срок выполнения всего комплекса не был сорван. Возможна и иная постановка: какие дополнительные ресурсы и в какие работы следует вложить, чтобы общий срок выполнения комплекса не превышал заданный, а стоимость дополнительных ресурсов минимизировалась.

5. ТРАНСПОРТНАЯ ЗАДАЧА

5.1. ПОСТАНОВКА ТРАНСПОРТНОЙ ЗАДАЧИ ПО КРИТЕРИЮ СТОИМОСТИ В МАТРИЧНОЙ ФОРМЕ

Транспортная задача (ТЗ) формулируется следующим образом. В m пунктах отправления $A_1, \dots, A_m$ сосредоточен однородный груз в количествах соответственно $a_1, \dots, a_m$ единиц. Имеющийся груз необходимо доставить потребителям $B_1, \dots, B_n$, спрос которых выражается величинами $b_1, \dots, b_n$ единиц. Известна стоимость c_{ij} перевозки единицы груза из i-го $(i = \overline{1,m})$ пункта отправления в j-й $(j = \overline{1,n})$ пункт назначения. Требуется составить план перевозок, который полностью удовлетворяет спрос потребителей в грузе, и при этом суммарные транспортные издержки минимизируются.

Для построения экономико-математической модели ТЗ рассмотрим матрицу

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix},$$

где x_{ij} $(i = \overline{1,m};\ j = \overline{1,n})$ обозначает количество единиц груза, которое необходимо доставить из i-го пункта отправления в j-й пункт назначения.

Матрицу $X = [x_{ij}]_{m \times n}$ будем называть *матрицей перевозок*. Предполагается, что все $x_{ij} \geqslant 0$. Удельные транспортные издержки (расходы) запишем в форме матрицы $C = [c_{ij}]_{m \times n}$ и назовем ее *матрицей тарифов*.

Для наглядности условия ТЗ можно представить таблицей (табл. 5.1), которую будем называть *распределительной*. Распределительную таблицу называют иногда *табличной* или *матричной моделью* ТЗ.

Экономико-математическая модель ТЗ должна отражать все условия и цель задачи в математической форме. Так, переменные x_{ij} $(i = \overline{1,m};\ j = \overline{1,n})$ должны удовлетворять ограничениям по запасам, потребностям и условиям неотрицательности. В математической форме эти условия можно записать так:

$$\left.\begin{aligned}\sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1,m}),\\ \sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1,n}),\end{aligned}\right\} \tag{5.1}$$

$$x_{ij} \geqslant 0 \quad (i = \overline{1,m};\ j = \overline{1,n}). \tag{5.2}$$

Таблица 5.1

Поставщик	Потребитель				Запас груза a_i
	B_1	B_2	...	B_n	
	Затраты на перевозку 1 ед. груза, доставка				
A_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	...	c_{2n} x_{2n}	a_2
...	...	...	...	...	...
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	...	c_{mn} x_{mn}	a_m
Потребность в грузе b_j	b_1	b_2	...	b_n	

Цель ТЗ — минимизировать общие затраты на реализацию плана перевозок, которые можно представить функцией

$$\begin{gathered} f = c_{11}x_{11} + c_{12}x_{12} + \ldots + c_{1n}x_{1n} + \ldots + c_{m1}x_{m1} + \\ + c_{m2}x_{m2} + \ldots + c_{mn}x_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}. \end{gathered} \tag{5.3}$$

Итак, математически ТЗ ставится так. Даны система ограничений (5.1) при условии (5.2) и линейная функция (5.3). Требуется среди множества решений системы (5.1) найти такое неотрицательное решение, при котором линейная функция (5.3) принимает минимальное значение (минимизируется).

Будем называть план перевозок $X = [x_{ij}]_{m\times n}$ *допустимым*, если он удовлетворяет ограничениям (5.1) и (5.2).

Допустимый план перевозок, доставляющий минимум целевой функции (5.3), называется *оптимальным*.

Теорема 5.1 **(о существовании допустимого плана).** *Для того чтобы ТЗ имела допустимые планы, необходимо и достаточно выполнение равенства*

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{5.4}$$

Доказательство. Просуммировав в распределительной таблице 5.1 элементы x_{ij} раздельно по индексам i и j, получим:

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1,n}), \qquad \sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1,m}).$$

Очевидно, что суммируются все элементы x_{ij} как по строкам, так и по столбцам, различие лишь в перестановке этих элементов. Однако от перестановки слагаемых сумма не меняется, поэтому равенство $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$ является необходимым условием разрешимости ТЗ.

Для доказательства достаточности условия (5.4) покажем, что если это условие выполняется, то всегда имеется допустимый план. Обозначим $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j = A$. Переменные x_{ij} выразим через данные задачи следующим образом:

$$x_{ij} = a_i b_j / A. \tag{5.5}$$

Поскольку $a_i \geqslant 0$, $b_j \geqslant 0$, то $A > 0$, а поэтому и $x_{ij} \geqslant 0$ $(i = \overline{1,m};\ j = \overline{1,n})$.

Покажем, что переменные (5.5) составляют допустимый план. Этот набор неотрицательных чисел будет составлять допустимый план тогда, когда он удовлетворяет системе ограничений (5.1). Просуммируем равенства (5.5) по индексу i. Получим

$$\sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} \frac{a_i b_j}{A}. \tag{5.6}$$

Поскольку индекс j не зависит от индекса i, в равенстве (5.6) вынесем b_j/A за знак суммы и получим

$$\sum_{i=1}^{m} x_{ij} = \frac{b_j}{A}\sum_{i=1}^{m} a_i = b_j \geqslant 0 \quad (j = \overline{1,n}).$$

Аналогично, суммируя равенства (5.5) по индексу j, получаем

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} \frac{a_i b_j}{A} = \frac{a_i}{A}\sum_{j=1}^{n} b_j = a_i \quad (i = \overline{1,m}).$$

Следовательно, набор чисел $x_{ij} = a_i b_j/A$ удовлетворяет системе ограничений задачи, а потому является допустимым планом. $\triangle$

5.2. ЗАКРЫТАЯ И ОТКРЫТАЯ МОДЕЛИ ТРАНСПОРТНОЙ ЗАДАЧИ

Модель ТЗ называют *закрытой*, если суммарный объем груза, имеющегося у поставщиков, равен суммарному спросу потребителей, т. е. выполняется равенство

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j.$$

Если для ТЗ выполняется одно из условий:

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j, \quad \sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j,$$

то модель задачи называют *открытой*.

Для разрешимости ТЗ с открытой моделью необходимо преобразовать ее в закрытую. Так, при выполнении первого условия необходимо ввести фиктивный $(n+1)$-й пункт назначения B_{n+1}, т. е. в матрице задачи предусматривается дополнительный столбец. Спрос фиктивного потребителя полагают равным небалансу, т. е. $b_{n+1} = \sum\limits_{i=1}^{m} a_i - \sum\limits_{j=1}^{n} b_j$, а все тарифы — одинаковыми, чаще всего равными нулю, т. е. $c_{i,n+1} = 0$ $(i = \overline{1,m})$.

Аналогично при выполнении второго условия вводится фиктивный поставщик A_{m+1}, запас груза у которого равен $a_{m+1} = \sum\limits_{j=1}^{n} b_j - \sum\limits_{i=1}^{m} a_i$, а тарифы дополнительной строки распределительной таблицы равны нулю, т. е. $c_{m+1,j} = 0$ $(j = \overline{1,n})$.

При преобразовании открытой задачи в закрытую целевая функция не меняется, так как все слагаемые, соответствующие дополнительным перевозкам, равны нулю.

Теорема 5.2 **(о ранге матрицы).** *Ранг матрицы A транспортной задачи на единицу меньше числа уравнений: $r(A) = m + n - 1$.*

Доказательство. Матрица системы ограничений (5.1) имеет вид

$$A = \begin{bmatrix} 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \multicolumn{13}{c}{\dotfill} \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 1 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & 1 & \ldots & 0 & \ldots & 0 & 1 & \ldots & 0 \\ \multicolumn{13}{c}{\dotfill} \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 1 & \ldots & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

В каждом столбце матрицы A содержатся только два элемента, равных единице, остальные элементы равны нулю. При этом, если сложить первые m строк матрицы, получим строку, элементами которой будут единицы. Этот же результат получаем, если сложить последние n строк. Обозначая i-ю строку через $\mathbf{p}_i$, получаем

$$\mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_m = \mathbf{p}_{m+1} + \mathbf{p}_{m+2} + \ldots + \mathbf{p}_{m+n}.$$

Отсюда видно, что любая строка есть линейная комбинация остальных строк, например

$$\mathbf{p}_1 = \mathbf{p}_{m+1} + \mathbf{p}_{m+2} + \ldots + \mathbf{p}_{m+n} - (\mathbf{p}_2 + \ldots + \mathbf{p}_m).$$

Значит, не меняя ранга матрицы A, можно вычеркнуть, например, последнюю строку. Минор $(m + n - 1)$-го порядка получившейся матрицы, составленный из столбцов коэффициентов при $x_{1n}, x_{2n}, \dots, x_{mn}, x_{11}, x_{12}, \dots, x_{1,n-1}$, будет отличен от нуля, что и доказывает теорему. $\triangle$

5.3. ПОСТРОЕНИЕ ИСХОДНОГО ОПОРНОГО ПЛАНА

Построение опорных планов, а также преобразование их будем производить непосредственно в распределительной таблице. Если в плане перевозок переменная x_{ik} равна некоторому числу $a \neq 0$, то это число записываем в соответствующую клетку $(i;k)$ и считаем ее занятой (или базисной), если же $x_{ik} = 0$, то клетку $(i;k)$ оставляем свободной. При этом число занятых опорным планом клеток в соответствии с теоремой 5.2 должно быть равно $m+n-1$, а остальные $mn-(m+n-1)=(m-1)(n-1)$ клеток будут свободными.

Рассмотрим *правило "северо-западного угла"*. Суть его состоит в следующем. Пользуясь табл. 5.1, будем распределять груз, начиная с загрузки левой верхней, условно называемой северо-западной, клетки (1; 1), двигаясь затем от нее по строке вправо или по столбцу вниз. В клетку (1; 1) занесем меньшее из чисел a_1, b_1, т. е. $x_{11} = \min(a_1, b_1)$. Если $a_1 > b_1$, то $x_{11} = b_1$ и первый потребитель B_1 будет полностью удовлетворен. В дальнейшем первый столбец таблицы в расчет не принимается; в нем переменные $x_{i1} = 0$ для $i = \overline{2,m}$.

Двигаясь вправо по первой строке таблицы, заносим в соседнюю клетку (1; 2) меньшее из чисел $a_1 - b_1$ и b_2, т. е. $x_{12} = \min(a_1 - b_1, b_2)$. Если $a_1 - b_1 < b_2$, то запасы первого поставщика исчерпаны и первая строка таблицы в дальнейшем в расчет не принимается. Переходим к аналогичному распределению запаса груза второго поставщика.

Если $b_1 > a_1$, то $x_{11} = \min(a_1, b_1) = a_1$. При этом запас первого поставщика будет исчерпан, а потому $x_{1k} = 0$ для $k = \overline{2,n}$. Первая строка из дальнейшего рассмотрения исключается.

Переходим к распределению запасов второго поставщика. В клетку (2; 1) заносим наименьшее из чисел a_2 и $b_1 - a_1$. Заполнив таким образом клетку (1; 2) или (2; 1), переходим к загрузке следующей клетки по второй строке либо по второму столбцу. Процесс распределения по второй, третьей и последующим строкам (столбцам) производится аналогично распределению по первой строке или первому столбцу до тех пор, пока не исчерпаются ресурсы. Последней заполняется клетка $(m; n)$.

Проиллюстрируем правило "северо-западного угла" на примере.

Пример 5.1. Составить план перевозок зерна из районов A_1, A_2, A_3 и A_4 республики, в которых запасы составляют соответственно 800, 700, 1000 и 500 тыс. ц, на три элеватора B_1, B_2 и B_3 мощностью 1000, 1100 и 900 тыс. ц. Затраты на перевозку 1 ц зерна из районов на элеваторы приведены в табл. 5.2.

Таблица 5.2

Районы	Элеваторы			Запас a_i, ц
	B_1	B_2	B_3	
	Затраты на перевозку 1 ц зерна, ден. ед.			
A_1	3	5	6	800
A_2	7	2	4	700
A_3	4	3	5	1000
A_4	6	4	7	500
Мощность b_j элеватора, ц	1000	1100	900	3000

Таблица 5.3

	B_1	B_2	B_3	a_i
A_1	3 800	5	6	800
A_2	7 200	2 500	4	700
A_3	4	3 600	5 400	1000
A_4	6	4	7 500	500
b_j	1000	1100	900	

Решение. Установим характер задачи. Сравнивая $\sum_{i=1}^{4} a_i = 800+$

$+700+1000+500=3000$ и $\sum\limits_{j=1}^{3} b_j = 1000+1100+900=3000$, заключаем, что данная ТЗ обладает закрытой моделью. В клетку (1; 1) табл. 5.3 помещаем $x_{11} = \min(800, 1000) = 800$ тыс. ц зерна. Весь запас зерна района A_1 отгружен на элеватор B_1. Недостающее количество зерна на элеватор B_1 поставляется из района A_2: в клетку (2; 1) помещаем $x_{21} = \min(700, 1000-800) = 200$ тыс. ц. В этом случае мощность элеватора B_1 будет полностью использована. Остаток зерна района A_2 отправляем на элеватор B_2: в клетку (2; 2) помещаем $x_{22} = \min(700-200, 1100) = 500$ тыс. ц зерна. Запас зерна района A_2 исчерпан, и переходим к перевозке зерна района A_3. В клетку (3; 2) помещаем $x_{32} = \min(1000, 1100-500) = 600$ тыс. ц зерна. Мощность элеватора B_2 полностью использована. Поставку зерна производим на элеватор B_3. В клетку (3; 3) помещаем $x_{33} = \min(900, 1000-600) = 400$ тыс. ц зерна. Отгрузка зерна из района A_3 полностью завершена. Производим отгрузку зерна из района A_4. В клетку (4; 3) помещаем $x_{43} = \min(500, 900-400) = 500$ тыс. ц зерна. В результате полной отгрузки зерна на элеваторы получили план перевозок X_1 (табл. 5.3), по которому на элеватор B_1 следует поставить 800 тыс. ц зерна из района A_1 и 200 тыс. ц из района A_2; на элеватор B_2 — 500 тыс. ц из района A_2 и 600 тыс. ц из района A_3; на элеватор B_3 — 400 тыс. ц из района A_3 и 500 тыс. ц из района A_4. Суммарные расходы на перевозку зерна составят $f(X_1) = 3\cdot 800 + 7\cdot 200 + 2\cdot 500 + 3\cdot 600 + 5\cdot 400 + 7\cdot 500 = 12\,100$ ден. ед.

Рассмотрим *правило "минимального элемента"*. Суть его состоит в следующем. Просматриваются тарифы табл. 5.1 и в первую очередь заполняется клетка с минимальным значением тарифа. При этом в клетку записывается максимально возможное значение поставки. Затем из рассмотрения исключают строку, соответствующую поставщику, запасы которого полностью израсходованы, или столбец, соответствующий потребителю, спрос которого полностью удовлетворен. После этого из оставшихся клеток таблицы снова выбирают клетку с наименьшим тарифом. Процесс распределения заканчивается, когда все запасы поставщиков исчерпаны, а спрос потребителей полностью удовлетворен. В результате получаем опорный план, который должен содержать $m+n-1$ загруженных клеток. В процессе заполнения таблицы могут быть одновременно исключены строка и столбец. Так бывает, когда полностью исчерпывается запас груза и полностью удовлетворяется спрос (вырожденная задача). В этом случае в свободные клетки надо записать число 0 — "*нуль-загрузку*", условно считая такую клетку занятой. Однако число 0 записывается в те свободные клетки, которые не образуют циклов с ранее занятыми клетками.

Проиллюстрируем правило "минимального элемента" для транспортной задачи, представленной табл. 5.2, и сравним значения целевых функций для планов, полученных по правилам "северо-западного угла" и "минимального элемента".

Просматривая табл. 5.2, замечаем, что наименьшие затраты на перевозку зерна соответствуют маршруту $A_2 - B_2$, поэтому в клетку (2; 2) табл. 5.4 помещаем $x_{22} = \min(700, 1100) = 700$ тыс. ц зерна. В этом случае вторая строка таблицы в дальнейшем в расчет не принимается, так как запас зерна в районе A_2 полностью доставлен на элеватор B_2. Просматриваем оставшиеся клетки таблицы. Наименьшие тарифы имеют клетки (1; 1), (3; 2): $c_{11} = c_{32} = 3$. В клетку (1; 1) помещаем $x_{11} = \min(800, 1000) = 800$ тыс. ц, а в клетку (3; 2) — $x_{32} = \min(1000, 1100 - 700) = 400$ тыс. ц.

Таблица 5.4

	B_1	B_2	B_3	a_i
A_1	3 800	5	6	800
A_2	7	2 700	4	700
A_3	4 200	3 400	5 400	1000
A_4	6	4	7 500	500
b_j	1000	1100	900	

Далее по величине тарифа следует загружать клетки (3; 1), (4; 2), (2; 3), так как $c_{31} = c_{42} = c_{23} = 4$. Однако в результате загрузки клеток (1; 1), (2; 2), (3; 2) запас зерна в районах A_1 и A_2 и частично в районе A_3 исчерпан. Мощность элеватора B_2 полностью использована, а мощность элеватора B_1 использована на 800 тыс. ц. Поэтому помещаем необходимое количество зерна в клетку (3; 1): $x_{31} = \min(1000 - 400, 1000 - 800) = \min(600, 200) = 200$ тыс. ц. После этого мощность элеватора B_1 полностью использована. Остался элеватор B_3, который может принять зерно из районов A_3 и A_4. В районе A_3 осталось зерна $x_{33} = 1000 - 400 - 200 = 400$ тыс. ц, а в районе A_4 — $a_4 = 500$ тыс. ц. В клетки (3; 3); (4; 3) помещаем необходимые количества зерна: $x_{33} = 400$ тыс. ц, $x_{43} = 500$ тыс. ц. В результате полного распределения зерна получаем план X_2, для которого значение целевой функции $f(X_2) = 3 \cdot 800 + 2 \cdot 700 + 4 \cdot 200 + 3 \cdot 400 + 5 \cdot 400 + 7 \cdot 500 = 11\,300$ ден. ед.

Сравнивая значения целевых функций для планов X_1 и X_2, полученных по правилам "северо-западного угла" и "минимального элемента", замечаем, что транспортные расходы по плану X_2 на перевозку зерна из районов на элеваторы меньше на 800 ден. ед.

Рассмотрим *метод Фогеля.* Суть его состоит в следующем. В табл. 5.1 по строкам и столбцам определяется разность между двумя наименьшими тарифами. Отмечается наибольшая разность знаком □. Далее в строке (столбце) с наибольшей разностью заполняется клетка с наименьшим тарифом. Строки (столбцы) с нулевым остатком груза в дальнейшем в расчет не принимаются. На каждом этапе загружается только одна клетка. Распределение груза производится, как и ранее.

Проиллюстрируем правило Фогеля (табл. 5.5) задачей, представленной в табл. 5.2.

Таблица 5.5

		B_1	B_2	B_3	a_i	Этапы				
						№ 1	№ 2	№ 3	№ 4	№ 5
A_1		3 800_1	5	6	800	$\boxed{2}$	—	—	—	
A_2		7	2 700_2	4	700	2	$\boxed{2}$	—	—	—
A_3		4 200_4	3	5 800_5	1000	1	1	1	1	—
A_4		6	4 400_3	7 100_6	500	2	2	$\boxed{2}$	1	—
b_j		100	1100	900						
Этапы	№ 1	1	1	1						
	№ 2	2	1	1						
	№ 3	2	1	2						
	№ 4	$\boxed{2}$	—	2						
	№ 5	—	—	$\boxed{2}$						

Для большей наглядности в табл. 5.5 поставки x_{ij} снабжены индексами, указывающими последовательность загрузки клеток таблицы.

Получен исходный опорный план X_3, для которого значение целевой функции $f(X_3) = 3 \cdot 800 + 2 \cdot 700 + 4 \cdot 200 + 5 \cdot 800 + 4 \cdot 400 + 7 \cdot 100 =$ $= 10\,900$ ден. ед.

Среди полученных опорных планов X_1, X_2, X_3 с различными значениями целевой функции лучшим является X_3. Будет ли он оптимальным? Чтобы ответить на этот вопрос, необходимо знать признак оптимальности.

Теорема 5.3. *Если план* $X^* = \left[x_{ij}^*\right]_{m\times n}$ *транспортной задачи является оптимальным, то ему соответствует система из* $m+n$ *чисел* u_i^*, v_j^*, *удовлетворяющих условиям* $u_i^* + v_j^* = c_{ij}$ *для* $x_{ij}^* > 0$ *и* $u_i^* + v_j^* \leqslant c_{ij}$ *для* $x_{ij}^* = 0$ $(i = \overline{1,m};$ $j = \overline{1,n})$.

Числа u_i^* и v_j^* называются *потенциалами* соответственно i-го поставщика и j-го потребителя.

Доказательство. Транспортную задачу (5.1) — (5.3) можно рассматривать как двойственную задачу к некоторой исходной задаче, решаемой на максимум. Построим эту задачу. Так, если i-му ограничению двойственной задачи $x_{i1} + x_{i2} + \ldots + x_{in} = a_i$ в исходной задаче соответствует переменная u_i $(i = \overline{1,m})$, а j-му ограничению $x_{1j} + x_{2j} + \ldots + x_{mj} = b_j$ — переменная v_j $(j = \overline{1,n})$, то исходной будет задача: найти максимальное значение функции

$$f = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j$$

при ограничениях

$$u_i + v_j \leqslant c_{ij} \quad (i = \overline{1,m};\ j = \overline{1,n}).$$

Оптимальным для двойственной задачи является план X^*, а для исходной $\mathbf{y}^* = (u_i^*; v_j^*)$. Для пары взаимно двойственных задач на основании первой теоремы двойственности имеет место равенство $\tilde{f}_{\min} = f_{\max}$, а по второй теореме двойственности выполняются условия $u_i^* + v_j^* = c_{ij}$ для $x_{ij}^* > 0$, $u_i^* + v_j^* \leqslant c_{ij}$ для $x_{ij}^* = 0$ $(i = \overline{1,m};\ j = \overline{1,n})$. $\triangle$

Из теоремы следует, что для оптимального плана ТЗ необходимо выполнение условий:

1) каждой занятой клетке в распределительной таблице соответствует сумма потенциалов, равная тарифу этой клетки, т. е. $u_i + v_j = c_{ij}$;

2) каждой свободной клетке соответствует сумма потенциалов, не превышающая тарифа этой клетки, т. е. $u_i + v_j \leqslant c_{ij}$.

Доказанная теорема носит название *теоремы о потенциалах*. На ней основан специальный метод решения ТЗ, названный *методом потенциалов.*

5.4. МЕТОД ПОТЕНЦИАЛОВ

В соответствии с введенным понятием потенциалов и с учетом связей между моделями двойственных задач каждому поставщику (ограничению по запасам) поставим в соответствие потенциал u_i $(i = \overline{1,m})$, а каждому потребителю (ограничению по спросу) — потенциал v_j $(j = \overline{1,n})$.

Согласно теореме о потенциалах, каждой занятой клетке будет соответствовать уравнение $u_i + v_j = c_{ij}$. Так как всех занятых клеток должно быть $m+n-1$, т. е. на единицу меньше числа потенциалов, то для определения чисел u_i, v_j необходимо решить систему из $m+n-1$ уравнений $u_i + v_j = c_{ij}$ с $m+n$ неизвестными. Система неопределенная, и, чтобы найти частные решения, одному из потенциалов придаем произвольное числовое значение, тогда остальные потенциалы определяются однозначно. Для облегчения расчетов одному из потенциалов придают обычно значение, равное нулю.

Для исследования плана на оптимальность по каждой свободной клетке проверяется условие $u_i + v_j \leqslant c_{ij}$. Если хотя бы одна свободная клетка не удовлетворяет данному условию, то опорный план не является оптимальным, его можно улучшить за счет загрузки этой клетки. Если таких клеток несколько, то наиболее перспективной для загрузки является клетка, для которой разность (оценка) между тарифом клетки и суммой потенциалов наименьшая, т. е.

$$s_{ij} = c_{ij} - (u_i + v_j) < 0.$$

Например, для клеток $(i;k)$ и $(i;t)$ имеем оценки: $s_{ik} = -5$, $s_{it} = -10$. Здесь наиболее потенциальной является клетка $(i;t)$. Экономически оценка показывает, на сколько денежных единиц уменьшатся транспортные издержки от загрузки данной клетки единицей груза. Так, от загрузки клетки $(i;k)$ единицей груза транспортные издержки уменьшатся на $\Delta f_1 = -5$ ден. ед., а от загрузки единицей груза клетки $(i;t)$ — на $\Delta f_2 = -10$ ден. ед. Эффективность плана от загрузки потенциальной клетки грузом в λ единиц составит $\Delta f = s_{ij}\lambda$ ден. ед.

Если для всех свободных клеток оценки $s_{ij} \geqslant 0$, то опорный план перевозок является оптимальным.

Итак, если для опорного плана перевозок указанное условие оптимальности не выполняется, то за счет загрузки свободной клетки с отрицательной оценкой план перевозок улучшается. Для наиболее перспективной свободной клетки строится замкнутый цикл с вершинами в загруженных клетках. Вершинам этого цикла условно приписываются знаки: свободной клетке — плюс, следующей по часовой или против часовой стрелки занятой клетке — минус, следующей — снова плюс и т. д. Из поставок в клетках цикла с "отрицательными" вершинами выбирается наименьшее количество λ груза, которое и перемещается по клеткам этого цикла: прибавляется к поставкам в положительных вершинах и вычитается из поставок в отрицательных вершинах, в результате чего баланс цикла не нарушится (рис. 5.1, *а*, *б*).

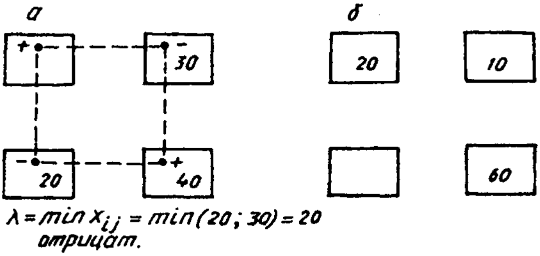


Рис. 5.1

В общем случае цикл представляет собой замкнутую ломаную линию, состоящую из звеньев, пересекающихся под прямым углом. Каждое звено соединяет две клетки строки (столбца). Цикл включает одну свободную клетку, остальные клетки цикла заняты. В цикле всегда четное число клеток. Для свободной клетки всегда можно построить единственный цикл. Если из занятых клеток образуется цикл, то план перевозок не является опорным. Цикл строится лишь для свободной клетки.

Сформулируем алгоритм решения ТЗ методом потенциалов:

1) построить опорный план по одному из правил;

2) вычислить потенциалы поставщиков и потребителей u_i и v_j $(i = \overline{1,m};\ j = \overline{1,n})$, решив систему уравнений вида $u_i + v_j = c_{ij}$;

3) вычислить оценки s_{ij} для всех свободных клеток по формуле $s_{ij} = c_{ij} - (u_i + v_j)$. Если все $s_{ij} \geqslant 0$, то полученный план является оптимальным. При этом если все $s_{ij} > 0$, то полученный оптимальный план единственный. В случае, если хотя бы одна оценка $s_{ij} = 0$, имеем бесчисленное множество оптимальных планов с одним и тем же значением целевой функции.

Пример 5.2. С трех складов A_1, A_2, A_3 необходимо доставить овощи в пять торговых точек B_1, B_2, B_3, B_4, B_5. Требуется закрепить склады за торговыми точками так, чтобы общая сумма затрат на перевозку была минимальной. Числовые данные задачи представлены в табл. 5.6.

Таблица 5.6

Склады	Торговые точки					Объем вывоза, т
	B_1	B_2	B_3	B_4	B_5	
	Стоимость перевозки 1 т груза, ден. ед.					
A_1	7	3	5	4	2	40
A_2	6	2	3	1	7	150
A_3	3	5	2	6	4	100
Объем вывоза, т	20	80	90	60	40	290

Решение. Исходное опорное решение получим, например, по правилу "минимального элемента" (табл. 5.7).

Получен опорный вырожденный план. Число занятых клеток не удовлетворяет условию $m+n-1 = 3+5-1 = 7$. В одну из свободных клеток помещаем число 0 и считаем такую клетку занятой. Поместим число 0, например, в клетку (1; 2) с наименьшим тарифом. План будет опорным, так как из занятых клеток не образуется циклов.

Для определения потенциалов составляем уравнения: $u_1 + v_2 = 3$, $u_1 + v_5 = 2$, $u_2 + v_1 = 6$, $u_2 + v_2 = 2$, $u_2 + v_4 = 1$, $u_3 + v_1 = 3$, $u_3 + v_3 = 2$. Положим, например, $u_1 = 0$, тогда $u_2 = -1$, $u_3 = -4$, $v_1 = 7$, $v_2 = 3$, $v_3 = 6$, $v_4 = 2$, $v_5 = 2$.

Потенциалы проставлены в табл. 5.7. Их можно вычислять и непосредственно в таблице, не выписывая систему уравнений. Ведь если известны потенциал и тариф занятой клетки, то из соотношения $u_i + v_j =$

Таблица 5.7

	20	80	90	60	40	
40	7	3 0	5	4	2 40	$u_1 = 0$
150	6 10	2 80	3 +	1 60	7	$u_2 = -1$
100	3 + 10	5	2 90	6	4	$u_3 = -4$
	$v_1 = 7$	$v_2 = 3$	$v_3 = 6$	$v_4 = 2$	$v_5 = 2$	

$= c_{ij}$ легко определить неизвестный потенциал (из суммы вычесть известное слагаемое, получится неизвестное слагаемое). Роль суммы в данном равенстве играет тариф c_{ij}.

Определим оценки свободных клеток:

$s_{11} = 7 - (0 + 7) = 0,$ $\qquad s_{13} = 5 - (0 + 6) = -1 < 0,$

$s_{14} = 4 - (0 + 2) = 2 > 0,$ $\qquad s_{23} = 3 - (-1 + 6) = -2 < 0,$

$s_{25} = 7 - (-1 + 2) = 6 > 0,$ $\qquad s_{32} = 5 - (-4 + 3) = 6 > 0,$

$s_{34} = 6 - (-4 + 2) = 8 > 0,$ $\qquad s_{35} = 4 - (-4 + 2) = 6 > 0.$

Таблица 5.8

	20	80	90	60	40	
40	7	3 0	5	4	2 40	$u_1 = 0$
150	6	2 80	3 10	1 60	7	$u_2 = -1$
100	3 20	5	2 80	6	4	$u_3 = -2$
	$v_1 = 5$	$v_2 = 3$	$v_3 = 4$	$v_4 = 2$	$v_5 = 2$	

Перспективными являются клетки (1; 3) и (2; 3) с оценками $s_{13} = -1$ и $s_{23} = -2$. Наиболее потенциальной является клетка (2; 3). Строим для клетки (2; 3) цикл непосредственно в таблице. В цикл войдут клетки (2; 3), (2; 1), (3; 1), (3; 3). Наименьшее количество груза, стоящее в вершинах цикла с отрицательным знаком, $\lambda = \min(10, 90) = 10$. В результате смещения λ по циклу получим новый план (табл. 5.8).

Для нового плана определяем новые потенциалы и оценки свободных клеток: $s_{11} = 2 > 0$, $s_{13} = 1 > 0$, $s_{14} = 2 > 0$, $s_{21} = 2 > 0$, $s_{25} = 6 > 0$, $s_{32} = 4 > 0$, $s_{34} = 6 > 0$, $s_{35} = 4 > 0$.

Оценки всех свободных клеток неотрицательны, следовательно, полученный план является оптимальным:

$$X^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 40 \\ 0 & 80 & 10 & 60 & 0 \\ 20 & 0 & 80 & 0 & 0 \end{bmatrix}.$$

Минимальные транспортные издержки для этого плана

$$\min f = f(X^*) = 2 \cdot 40 + 2 \cdot 80 + 3 \cdot 10 + 1 \cdot 60 + 3 \cdot 20 + 2 \cdot 80 = 520.$$

Согласно оптимальному плану, со склада A_1 нужно поставить 40 т овощей в торговую точку B_5, со склада A_2 — 80 т овощей в торговую точку B_2, 10 т в торговую точку B_3 и 60 т в торговую точку B_4, со склада A_3 — 20 т овощей в торговую точку B_1 и 80 т в торговую точку B_3.

5.5. РЕШЕНИЕ ТРАНСПОРТНОЙ ЗАДАЧИ С ОТКРЫТОЙ МОДЕЛЬЮ

Решение ТЗ проиллюстрируем для случая, когда $\sum\limits_i a_i > \sum\limits_j b_j$.

Пример 5.3. В трех хранилищах A_1, A_2, A_3 имеется соответственно 70, 90 и 50 т топлива. Требуется спланировать перевозку топлива четырем потребителям B_1, B_2, B_3, B_4, спрос которых равен соответственно 50, 70, 40 и 40 т, так, чтобы затраты на транспортировку были минимальны. Стоимость перевозки 1 т указана в табл. 5.9.

Решение. Поскольку запасы топлива в хранилищах превышают спрос потребителей, вводится фиктивный потребитель, спрос которого

$$b_5 = \sum_{i=1}^{3} a_i - \sum_{j=1}^{4} b_j = 210 - 200.$$

Все затраты для фиктивного потребителя $c_{i5} = 0$ $(i = \overline{1,3})$. После введения фиктивного потребителя открытая модель задачи преобразовалась в закрытую, а распределительная таблица принимает вид табл. 5.10.

Исходный опорный план получим, например, по правилу "минимального элемента". Так как наименьшими являются нулевые тарифы для

Таблица 5.9

Хранилища	Потребители				Запас топлива, т
	B_1	B_2	B_3	B_4	
	Стоимость перевозки 1 т топлива, ден. ед.				
A_1	5	2	3	6	70
A_2	4	3	5	7	90
A_3	2	4	1	5	50
Потребность в топливе, т	50	70	40	40	$210 > 200$

Таблица 5.10

Хранилища	Потребители					Запас топлива, т
	B_1	B_2	B_3	B_4	B_5	
	Стоимость перевозки 1 т топлива, ден. ед.					
A_1	5	2	3	6	0	70
A_2	4	3	5	7	0	90
A_3	2	4	1	5	0	50
Потребность в топливе, т	50	70	40	40	10	210

клеток (1; 5), (2; 5), (3; 5), то загрузим первой, например, клетку (1; 5), $x_{15} = 10$ т. Второй загружаем клетку (3; 3), $x_{33} = 40$ т. Далее загружаем клетки (1; 2), (3; 1), (2; 2), (2; 4), полагая $x_{12} = 60$ т, $x_{31} = 10$ т, $x_{22} = 10$ т, $x_{21} = 40$ т, $x_{24} = 10$ т (табл. 5.11).

В результате распределения топлива по потребителям получили невырожденный план: условие для занятых клеток $m+n-1=3+5-1=7$ выполняется.

Определив потенциалы (см. табл. 5.11), устанавливаем, что среди оценок свободных клеток одна отрицательная: $s_{25} = -1$. Следовательно, план перевозок можно улучшить за счет загрузки клетки (2; 5). Цикл для нее выделен штриховой линией в табл. 5.11.

Наименьшее количество топлива в отрицательных вершинах цикла равно 10 т. После смещения по циклу 10 т получаем новый план перевозок (табл. 5.12).

Полученный план является вырожденным. Поставим число 0, например, в клетку (2; 2).

Вычислив потенциалы, находим оценки свободных клеток: $s_{11} = 2 > 0$, $s_{13}=1>0$, $s_{14}=0$, $s_{15} = 1 > 0$, $s_{23} = 2 > 0$, $s_{32} = 3 > 0$, $s_{34} = 0$, $s_{35} = 2 > 0$.

Оценки всех свободных клеток $s_{ij} \geqslant 0$, следовательно, получен оптимальный план. Поскольку среди оценок имеются равные нулю, то за счет загрузки клеток (1; 4), (3; 4) можно получить новые планы, но

Таблица 5.11

	50	70	40	40	10	
70	5	2 + 60	3	6	0 10	$u_1 = 0$
90	4 40	3 10	5	7 40	0 +	$u_2 = 1$
50	2 10	4	1 40	5	0	$u_3 = -1$
	$v_1 = 3$	$v_2 = 2$	$v_3 = 2$	$v_4 = 6$	$v_5 = 0$	

Таблица 5.12

	50	70	40	40	10	
70	5	2 70	3	6	0	$u_1 = -1$
90	4 40	3 0	5	7 40	0 10	$u_2 = 0$
50	2 10	4	1 40	5	0	$u_3 = -2$
	$v_1 = 4$	$v_2 = 3$	$v_3 = 3$	$v_4 = 7$	$v_5 = 0$	

значение целевой функции не изменится. Это случай бесчисленного множества оптимальных планов.

Итак, в табл. 5.12 получили оптимальный план

$$X^* = \begin{bmatrix} 0 & 70 & 0 & 0 \\ 40 & 0 & 0 & 40 \\ 10 & 0 & 40 & 0 \end{bmatrix},$$

для которого значение целевой функции

$$f(X^*) = 2 \cdot 70 + 4 \cdot 40 + 7 \cdot 40 + 2 \cdot 10 + 1 \cdot 40 = 640.$$

Десять тонн топлива, находящегося в хранилище A_2, осталось нераспределенным.

6. ДИСКРЕТНОЕ ПРОГРАММИРОВАНИЕ

6.1. КЛАССИЧЕСКИЕ ЗАДАЧИ ЦЕЛОЧИСЛЕННОГО ПРОГРАММИРОВАНИЯ И КРАТКАЯ КЛАССИФИКАЦИЯ МЕТОДОВ ИХ РЕШЕНИЯ

Основные понятия. *Дискретное программирование* — раздел математического программирования, изучающий экстремальные задачи, в которых на искомые переменные налагается условие целочисленности, а область допустимых решений конечна. В экономике существует огромное количество задач с дискретной природой. Прежде всего это задачи с физической неделимостью многих факторов и объектов расчета. Например, нельзя построить 3,2 завода или поставить 1,6 автомобиля. Количество комплектов, число агрегатов, число типовых размеров предприятий, типовые мощности предприятий — все это вносит дискретность в оптимизационные расчеты. Дискретными являются задачи с логическими переменными, принимающими только два значения — нуль или единица (вариант отвергается или принимается).

Иногда дискретное программирование называется *целочисленным*. Этот термин некоторые математики считают неправильным, так как, строго говоря, дискретное — не обязательно целочисленное. Например, ряд вместимостей (в м^3) 1,3; 1,6; 1,9; ... — дискретный, но не целочисленный. Отсюда целочисленное программирование правильнее считать частным случаем дискретного. Обратимся к примерам.

Задача о контейнерных перевозках (задача о бомбардировщике или задача о рюкзаке). Эта задача формулируется так. Контейнер оборудован m отсеками вместимостью b_i $(i = \overline{1,m})$ для перевозки n видов продукции Π_j $(j = \overline{1,n})$. Виды продукции характеризуются свойством неделимости, т. е. их можно брать в количестве 0, 1, 2, ... единиц. Пусть, далее, a_{ij} — расход i-го отсека для перевозки единицы j-й продукции. Обозначим через c_j полезность единицы

j-й продукции (это может быть цена реализации, прибыль и др.). Требуется найти план $\{x_j\}$ (x_j — количество единиц j-й продукции, погруженной в контейнер) перевозки, при котором максимизируется общая полезность рейса. Модель задачи примет вид

$$\max Z = \sum_{j=1}^{n} c_j x_j$$

при ограничениях на вместимости отсеков

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1, m}),$$

условии неотрицательности

$$x_j \geqslant 0 \quad (j = \overline{1, n}),$$

условии целочисленности

$$x_j \text{ — целые} \quad (j = \overline{1, n}).$$

В частности, когда для перевозки имеется один отсек и каждый вид продукции (предмет) может быть взят или нет (т. е. $x_j \in \{0;\ 1\}$: $x_j = 1$, если предмет j-го вида берется, и $x_j = 0$ в противном случае), модель задачи принимает вид

$$\max Z = \sum_{j=1}^{n} c_j x_j; \tag{6.1}$$

$$\sum_{j=1}^{n} a_j x_j \leqslant b, \tag{6.2}$$

$$x_j \in \{0;\ 1\}. \tag{6.3}$$

Для модели (6.1) – (6.3) можно дать приближенное графическое решение задачи, которое приемлемо при условии, что a_j значительно меньше b. Для этого рассмотрим прямоугольную декартову систему координат. По оси ординат отложим значения c_j, а по оси абсцисс — a_j. Построим точки Π_j с координатами a_j, c_j. Для нахождения оптимального набора предметов Π_j будем вращать луч, начальное положение которого совпадает с осью c_j, вокруг начала координат по ходу часовой стрелки (рис. 6.1, a). По мере вращения луча будем находить сумму абсцисс точек $П_j(a_j; c_j)$, которые "заметает" луч. При выполнении условия

$$\sum_{j=1}^{n} a_j x_j \leqslant b < \sum_{j=1}^{n+1} a_j x_j$$

вращение луча прекращается. Все предметы $П_j$, которые попадают в "заметаемый" радиусом-вектором сектор до выпол-

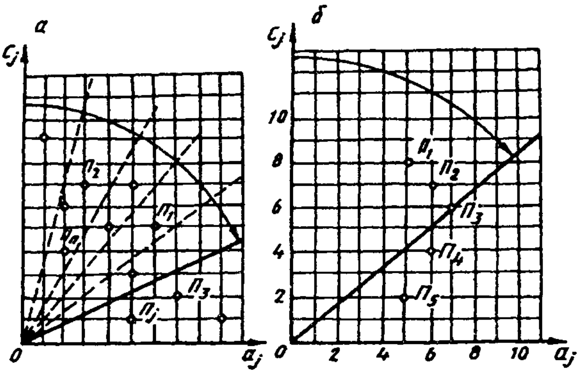


Рис. 6.1

нения этого условия, относятся к оптимальному набору. Можно обойтись и без графического представления условий. Для этого упорядочим отношения c_j/a_j по невозрастанию. Просматриваем их слева направо, следя за суммой a_j, и включаем соответствующий предмет в контейнер, пока в нем есть место. Если очередной предмет не вмещается в контейнер, то на этом не останавливаемся, откладываем предмет в сторону и переходим к следующему отношению, пока не пересмотрим все предметы либо не заполним контейнер. Такой подход в ряде случаев дает лучший результат, чем описанный выше графический способ.

Пример 6.1. Решить задачу дискретного программирования

$$\max Z = 8x_1 + 7x_2 + 6x_3 + 4x_4 + 2x_5;$$

$$5x_1 + 6x_2 + 7x_3 + 6x_4 + 5x_5 \leqslant 23,$$

$$x_j \in \{0; 1\}.$$

Решение. На рис. 6.1, *б* построены точки $П_1(5;8)$, $П_2(6;7)$, $П_3(7;6)$, $П_4(6;4)$, $П_5(5;2)$. При изображенном на рисунке положении луча в образовавшийся сектор попали точки $П_1, П_2, П_3$, для которых $a_1 + a_2 + a_3 = 5 + 6 + 7 = 18 < 23$. Если же продолжить вращение луча и "замести" точку $П_4$, то $a_1 + a_2 + a_3 + a_4 = 5 + 6 + 7 + 6 = 24 > 23$, т. е. ограничение задачи нарушается. Поэтому приближенным решением задачи при графическом способе будет набор предметов $П_1, П_2, П_3$, или аналитически $\mathbf{x}_1 = (1;1;1;0;0)$. При этом $Z = 21$.

Решим задачу другим способом. Последовательность отношений c_j/a_j имеет вид 8/5, 7/6, 6/7, 4/6, 2/5. Суммируем знаменатели a_j: $a_1 + a_2 = 5 + 6 = 11 < 23$. Следовательно, предметы $П_1$ и $П_2$ вмещаются в контейнер. Далее, $a_1 + a_2 + a_3 = 5 + 6 + 7 = 18 < 23$, поэтому предмет $П_3$ также вмещается в контейнер. Продолжаем суммирование: $a_1 + a_2 + a_3 + a_4 = 5 + 6 + 7 + 6 = 24 > 23$. Ограничение задачи нарушается, значит, предмет $П_4$ в контейнер не вмещается. Обращаемся к предмету $П_5$ и проверяем сумму: $a_1 + a_2 + a_3 + a_5 = 5 + 6 + 7 + 5 = 23$. Итак, предмет $П_5$ можно поместить в контейнер. Приходим к окончательному решению: $\mathbf{x}_2 = (1;1;1;0;1)$. При этом $Z = 23$. Понятно, что второе решение лучше.

Расчет погрешности не составляет труда. Если в задаче о контейнере отбросить условие целочисленности и заменить его условием $0 \leqslant x_j \leqslant 1$, то оптимальное значение целевой функции задачи будет оценкой сверху оптимального значения целевой функции $Z^*_ц$ задачи о контейнере, т. е.

$$z(\mathbf{x}) \leqslant Z^*_ц \leqslant [Z^*],$$

где $\mathbf{x}$ — приближенное решение задачи о контейнере; $[Z^*]$ — наибольшее целое число, не превосходящее Z^*. Отсюда погрешность приближенного решения не превосходит величины

$$\varepsilon = \frac{[Z^*] - z(\mathbf{x})}{z(\mathbf{x})}.$$

Для примера 6.1: $\mathbf{x}^* = (1;1;1;5/6;0;0)$, $[Z^*] = [z(\mathbf{x}^*)] = 24$, $\varepsilon = (24 - 23)/23 = 0,044$.

Задача о назначении (проблема выбора, задача о женихах и невестах). Эта задача уже рассматривалась в § 4.6. Она является исторически первой задачей дискретного программирования (опубликована венгерским математиком Е. Эгервари в 1932 г. как задача транспортного типа).

Имеется n исполнителей, которые могут выполнять n различных работ. Известна полезность c_{ij}, связанная с выполнением i-м исполнителем j-й работы $(i, j = \overline{1, n})$. Необходимо так назначить исполнителей на работы, чтобы добиться максимальной полезности при условии, что каждый исполнитель может быть назначен только на одну работу и за каждой работой должен быть закреплен только один исполнитель.

Для составления математической модели задачи обозначим через x_{ij} факт назначения или неназначения i-го исполнителя на j-ю работу. Так как количество исполнителей равно количеству работ и каждый из них может быть назначен только на одну работу, то x_{ij} должны принимать только два значения: 1 или 0. Такие переменные называют *булевыми*. Итак,

$$x_{ij} = \begin{cases} 1, & \text{если } i\text{-й исполнитель назначается на } j\text{-ю работу;} \\ 0 & \text{в противном случае.} \end{cases}$$

Приходим к задаче: найти план назначения x_{ij}, который максимизирует суммарную полезность назначений:

$$\max Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

при следующих ограничениях. Каждый исполнитель назначается только на одну работу:

$$\sum_{j=1}^{n} x_{ij} = 1 \quad (i = \overline{1, n}). \tag{6.4}$$

На каждую работу назначается только один исполнитель:

$$\sum_{i=1}^{n} x_{ij} = 1 \quad (j = \overline{1, n}). \tag{6.5}$$

Условия неотрицательности и целочисленности (булевости):

$$x_{ij} \geqslant 0, \quad x_{ij} \in \{0;\ 1\}. \tag{6.6}$$

Легко видеть, что задача о назначении — частный случай транспортной задачи при $a_i = 1$, $b_j = 1$. Однако с учетом специфики задачи для ее решения разработаны специальные, более эффективные, алгоритмы.

Задача о назначении имеет самое широкое применение. Например, при закреплении машин за маршрутами, распределении инструментов для обработки различных марок стали, рабочих или бригад и т. д. В каждом конкретном случае математическая модель задачи может иметь специфику. Например, в задаче распределения алгоритмов между вычислительными машинами на ВЦ число распределяемых алгоритмов, как правило, не равно числу машин. При назначении на должности для некоторых исполнителей существуют ограничения. Прежде всего отметим, что если задача о назначениях ставится при условии получения максимального эффекта, то ее сводят к задаче на минимум. Пусть дана матрица эффективности $C = [c_{ij}]$. В каждом столбце найдем максимальный элемент $l_j = \max_i c_{ij}$. Построим матрицу $C' = [l_j - c_{ij}]$. Задача

$$\max Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}, \quad \mathbf{x} \in \Omega,$$

где Ω — область допустимых решений системы (6.4) — (6.6), эквивалентна задаче

$$\min Z' = \sum_{i=1}^{n} \sum_{j=1}^{n} (l_j - c_{ij}) x_{ij}, \quad \mathbf{x} \in \Omega.$$

В самом деле,

$$Z' = \sum_{i=1}^{n} \sum_{j=1}^{n} (l_j - c_{ij}) x_{ij} = \sum_{j=1}^{n} l_j \left(\sum_{i=1}^{n} x_{ij} \right) - \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} = \sum_{j=1}^{n} l_j - Z.$$

Функция Z' достигает минимума при условии, что Z достигает максимума, $\mathbf{x} \in \Omega$.

Если в задаче о назначениях число исполнителей равно числу работ, то говорят о закрытой модели, в противном случае — об открытой модели задачи о назначениях. Если число m меньше числа исполнителей n $(m < n)$, то вводят $n - m$ фиктивных работ. Считается, что с назначением на фиктивные работы исполнителей не связаны затраты, т. е. соответствующие коэффициенты матрицы потерь равны нулю. В случае же $m > n$ вводят $m - n$ фиктивных исполнителей. Соответствующие элементы c_{ij} матрицы потерь полагают очень большими ("блокируют бесконечностью").

И еще одна деталь. Если по каким-либо причинам запрещается выполнение какой-либо работы каким-либо исполнителем, то и в этом случае соответствующую клетку "блокируют бесконечностью" (ставят бо́льшую стоимость M). Точное решение задачи о назначениях можно найти венгерским методом, методом динамического программирования и др. Существует также много приближенных методов. Хорошее приближение дает модификация метода Фогеля для определения опорного плана транспортной задачи.

Пример 6.2. При закреплении транспортных средств за маршрутами определена матрица прибылей

$$C = \begin{bmatrix} 13 & 10 & 9 & 12 \\ 7 & 11 & 9 & 5 \\ 12 & 13 & 15 & 10 \\ 9 & 14 & 10 & 8 \end{bmatrix}.$$

Найти план закрепления, максимизирующий суммарный эффект.

Решение. Прежде всего сведем задачу на максимум к задаче на минимум. Для этого в каждом столбце матрицы C найдем максимальный элемент и вычтем из него элемент соответствующего столбца. Получим матрицу C_1 задачи о назначениях:

$$C_1 = \begin{bmatrix} 0 & 4 & 6 & 0 \\ 6 & 3 & 6 & 7 \\ 1 & 1 & 0 & 2 \\ 4 & 0 & 5 & 4 \end{bmatrix},$$

которую будем решать на минимум:

$$\min Z' = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\max_i c_{ij} - c_{ij} \right) x_{ij}, \quad \mathbf{x} \in \Omega.$$

Для получения приближенного решения воспользуемся методом Фогеля для нахождения начального опорного плана транспортной задачи.

Таблица 6.1

	1	2	3	4	Δ_{i_1}	Δ_{i_2}	Δ_{i_3}
1	0	4	6	$\underline{0}$	0	0	—
2	$\underline{6}$	3	6	7	3	3	3
3	1	1	$\underline{0}$	2	1	—	—
4	4	$\underline{0}$	5	4	4	4	$\boxed{4}$
Δ_{j_1}	1	1	$\boxed{5}$	2	5	—	—
Δ_{j_2}	4	3	—	$\boxed{4}$	—	4	—
Δ_{j_3}	2	3	—	—	—	—	4

В каждом ряду матрицы C_1 находим минимальный и ближайший к нему элементы и их разность по абсолютной величине записываем в конце соответствующего ряда справа и снизу (табл. 6.1). Находим максимальную из этих разностей (число 5 заключено в рамку). В ряду, соответствующем максимальной разности, находим минимальный элемент ($c_{33} = 0$).

Мысленно вычеркиваем из матрицы (табл. 6.1) строку и столбец, соответствующие этому элементу. Фиксируем полученное назначение. Для нашего примера закрепляем третье транспортное средство за третьим маршрутом, это закрепление в табл. 6.1 подчеркнуто. С оставшейся матрицей поступаем аналогично предыдущему. Все вычисления сведены в табл. 6.1.

План назначения: (*1* — *4*), (*2* — *1*), (*3* — *3*), (*4* — *2*), $\max Z = c_{14} + c_{21} + c_{33} + c_{42} = 12 + 7 + 15 + 14 = 48$.

Задача коммивояжера (развозчика продукции или бродячего торговца). Коммивояжер должен посетить один, и только один, раз каждый из n городов и вернуться в исходный пункт. Его маршрут должен минимизировать суммарную длину пройденного пути.

Пусть $C = [c_{ij}]$ — матрица расстояний между городами. Неизвестные величины:

$$x_{ij} = \begin{cases} 1, & \text{если коммивояжер из города } i \text{ приезжает} \\ & \text{непосредственно в город } j; \\ 0 & \text{в противном случае.} \end{cases}$$

Модель задачи примет следующий вид:

$$\min Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}; \tag{6.7}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \quad (j = \overline{1, n}), \tag{6.8}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \quad (i = \overline{1, n}). \tag{6.9}$$

Система ограничений (6.8) обеспечивает построение маршрута, при котором коммивояжер въезжает в каждый город только один раз, а система ограничений (6.9) — маршрута, когда он выезжает из каждого города только один раз. К сожалению, эти ограничения недостаточны, так как среди допускаемых ими решений имеются маршруты, не образующие полный цикл, включающий все города. Устранение подциклов достигается при добавлении системы ограничений:

$$u_i - u_j + n x_{ij} \leqslant n - 1 \quad (i, j = \overline{1, n};\ i \neq j). \tag{6.10}$$

Ограничения (6.8) — (6.9) обусловливают задачу о назначении. Она отличается от задачи коммивояжера отсутствием требования цикличности пути. Аналитически требование цикличности записывается в виде неравенств (6.10), где переменные u могут принимать произвольные действительные значения. В самом деле, допустим обратное. Пусть имеется подцикл τ с числом ребер $k < n$, не проходящий через город 0. Складывая неравенства (6.10) при $x_{ij} = 1$ вдоль подцикла, получаем противоречивое неравенство $nk \leqslant (n-1)k$, так как все разности $u_i - u_j$ уничтожаются. Таким образом, условие (6.10) не допускает цикла, не проходящего через город 0. Вместе с тем покажем, что можно выбрать значения, при которых для цикла, проходящего через город 0 ($i = 0$), удовлетворяются соотношения (6.10). Для этого предположим, что $u_i = p$, если i посещается на p-м шаге ($p = 1, 2, \ldots$). Очевидно, что при этом $u_i - u_j \leqslant n - 1$ для любых i и j. Отсюда следует, что соотношения (6.10) для $x_{ij} = 0$ удовлетворяются. При $x_{ij} = 1$ они превращаются в равенства:

$$u_i - u_j + nx_{ij} = p - (p+1) + n \cdot 1 = n - 1.$$

Задача коммивояжера в математической постановке эквивалентна задаче упорядочения конечного числа работ на машине при учете потерь от переналадок. Во многих случаях длительность переналадок машины перед выполнением некоторой работы нельзя считать не зависящей от характера предшествующей работы. Иногда разброс длительностей переналадок становится основным критерием оценки расписания. Хорошей иллюстрацией задачи упорядочения с учетом издержек от переналадок может служить производство красок. Краски различного цвета можно получать на одном и том же оборудовании, которое должно очищаться при каждой перемене цвета изготавливаемой краски. Суммарная длительность очистки сильно зависит от последовательности перехода от одного цвета краски к другому. Возникает задача отыскания такой последовательности, в которой суммарные потери от переналадок минимальны. Для задачи упорядочения каждая работа соответствует городу, а длительность переналадки — расстоянию между городами. Матрица C, вообще говоря, может быть несимметричной, так как затраты, связанные с переналадкой с i-й работы на j-ю, как правило, отличны от затрат при переходе от j-й работы к i-й.

Существует несколько точных методов решения задачи коммивояжера: ветвей и границ, динамического программирования и др. Методом ветвей и границ с использованием современных ЭВМ можно решать задачи коммивояжера для $n \leqslant 40$, динамического программирования – для $n \leqslant 17$. В связи с малой эффективностью точных методов получили широкое применение эвристические. В настоящее время существует более ста приближенных методов решения задачи коммивояжера, среди которых наиболее прост *метод ближайшего соседа*. Он реализует требование включать в искомый замкнутый контур вершину, ближайшую к только что найденной. Алгоритм "ближайшего соседа" состоит в последовательном добавлении к начальной вершине следующей ближайшей к ней и т. д. Метод очень прост, однако степень приближения к оптимальному решению существенно зависит от выбора начальной вершины. Поэтому алгоритм целесообразно применять, начиная с каждой вершины, и затем выбрать замкнутый контур наименьшей длины. Отметим, что если ближайший сосед для некоторой вершины уже вошел в контур, то берется следующая по близости вершина и т. д.

Пример 6.3. Решить задачу коммивояжера для матрицы расстояний, представленной в табл. 6.2.

Таблица 6.2

i \ j	1	2	3	4	5
1	∞	35	45	20	11
2	9	∞	17	6	8
3	21	31	∞	2	11
4	30	15	40	∞	10
5	10	9	8	7	∞

Решение. Начнем с первой вершины. Ближайшей к ней является пятая ($\min\limits_j c_{ij} = 11$). Процесс нахождения минимального контура целесообразно сопровождать построением дерева ветвления (рис. 6.2). Если ближайшая вершина уже вошла в контур, то блокируем ее и переходим к следующей по степени близости. Ближайшей к пятой вершине является четвертая, к четвертой — пятая. Но она уже вошла в контур, поэтому блокируем ее и находим следующую по близости вершину — вторую, и т. д. Дерево ветвлений, начиная с первой вершины, представлено на рис. 6.2, *а*.

Для начала с других вершин соответствующие деревья представлены на рис. 6.2, *б* — *д*.

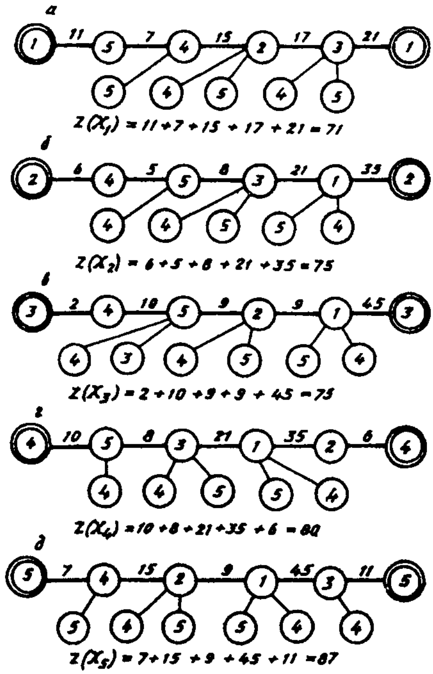


Рис. 6.2

В результате получаем

$$\min(z(x_1), z(x_2), z(x_3), z(x_4), z(x_5)) = \min(71, 75, 75, 80, 87) = 71.$$

Таким образом, минимальным контуром, найденным способом ближайшего соседа, является контур $\Gamma_1 = \langle 1 - 5 - 4 - 2 - 3 - 1 \rangle$, $z(\Gamma_1) = 71$. Дальше будет найдено точное решение методом ветвей и границ:

$$\Gamma^* = \langle 1 - 5 - 3 - 4 - 2 - 1 \rangle, \quad z(\Gamma^*) = 11 + 8 + 2 + 15 + 9 = 45.$$

Как видно, ошибка решения значительная.

На обширном статистическом материале показано, что с увеличением n ошибка решения убывает. Поэтому при $n \leqslant 40$ можно применять точные методы, при $n > 40$ — приближенные типа ближайшего соседа.

Классическим примером целочисленной задачи является обычная транспортная задача при целых значениях a_i и b_j.

Транспортная задача с фиксированными доплатами. Пусть имеется m пунктов производства с объемами a_i $(i = \overline{1, m})$ и n пунктов потребления с объемами b_j $(j = \overline{1, n})$. Требуется определить объемы поставок x_{ij} по коммуникациям (i, j), удовлетворяющие ограничениям транспортной задачи и минимизирующие целевую функцию

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij},$$

где

$$c_{ij} = \begin{cases} d_{ij} + c_{ij} x_{ij}, & x_{ij} > 0, \\ 0, & x_{ij} = 0. \end{cases}$$

В этой задаче $c_{ij} \geqslant 0$ — стоимость перевозки единицы груза от i-го поставщика j-му потребителю; $d_{ij} \geqslant 0$ — фиксированные доплаты, связанные с единовременными капиталовложениями на развитие производственно-комплектовочной базы, строительство подъездных путей и т. д. Это модель транспортной задачи с фиксированными доплатами. Она является нелинейной задачей с разрывной целевой функцией. Задачу с фиксированными доплатами можно свести к задаче частично целочисленного программирования. Введем переменную

$$\delta_{ij} = \begin{cases} 1, & \text{если } x_{ij} > 0, \\ 0, & \text{если } x_{ij} = 0. \end{cases}$$

Пусть, далее, $M_{ij} = \max\{a_i, b_j\}$. Тогда задача с фиксированными доплатами и следующая задача с булевыми переменными эквивалентны:

$$\min Z = \sum_{i=1}^{m} \sum_{j=1}^{n} (d_{ij} \delta_{ij} + c_{ij} x_{ij}),$$

$$\delta_{ij} = \begin{cases} 1, & \text{если } x_{ij} = M_{ij} \delta_{ij}, \\ 0, & \text{если } x_{ij} < M_{ij} \delta_{ij}, \end{cases}$$

$$\sum_{j=1}^{n} x_{ij} = a_i \quad (i = \overline{1, m}),$$
$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1, n}). \qquad (6.11)$$

Легко показать, что если в транспортной задаче с фиксированными доплатами все $d_{ij} = d = \text{const}$ и соответствующая ей задача без доплат с матрицей $[c_{ij}]$ не вырождена, то оптимальные решения обеих задач совпадают. Значения же обеих целевых функций будут различаться на постоянную величину $d(m + n - 1)$. В самом деле, так как в невырожденной задаче число переменных, отличных от нуля, равно $m + n - 1$, то

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(d_{ij}\delta_{ij} + c_{ij}x_{ij}) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + d(m + n - 1).$$

Если же d_{ij} в транспортной задаче с фиксированными доплатами различны, то ее оптимальное решение следует искать среди опорных решений соответствующей ей транспортной задачи без доплат с той же матрицей.

Для решения задачи с фиксированными доплатами может быть применен приближенный метод. Отбросим условие целочисленности для δ_{ij}. Примем $\delta_{ij} = x_{ij}/M_{ij}$. Получим обычную транспортную задачу, аппроксимирующую исходную:

$$\min Z = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{d_{ij}}{M_{ij}} + c_{ij}\right)x_{ij}, \quad \mathbf{x} \in \Omega. \qquad (6.12)$$

При ограничениях (6.11) решение задачи (6.12) определяет приближенное решение исходной задачи с фиксированными доплатами. В самом деле, если $\{x_{ij}^*\}$ — оптимальное решение аппроксимирующей задачи, то приближенное решение исходной определится так: $x_{ij} = \delta_{ij} = 0$, если $x_{ij}^* = 0$; $x_{ij} = x_{ij}^*$, $\delta_{ij} = 1$, если $x_{ij}^* > 0$.

Модели развития и размещения. Рассмотрим сначала одноэтапную статическую модель. Пусть в i-х пунктах могут быть построены предприятия с возможным набором мощностей и удельными затратами производства α_{ik} $(i = \overline{1, m};$ $k = \overline{1, k_i})$. Необходимо выбрать один из вариантов.

Обозначим через δ_{ik} переменную, принимающую значение 1, если в i-м пункте будет принят k-й вариант, и 0 в противном случае.

В простейшей модели развития и размещения территориальное расположение потребителей продукции не учитывается. Пусть b — общая потребность района в продукции. Получим модель задачи минимизации функции

$$Z = \sum_{i=1}^{m} \sum_{k=1}^{k_i} \alpha_{ik}\rho_{ik}\delta_{ik}$$

(ρ_{ik} — мощность предприятия в i-м пункте при k-м варианте размещения) при следующих ограничениях. Потребность экономического района в продукции должна быть удовлетворена полностью:

$$\sum_{i=1}^{m} \sum_{k=1}^{k_i} \alpha_{ik}\rho_{ik}\delta_{ik} \geqslant b.$$

Для пунктов K_i', где обязательно наличие предприятия,

$$\sum_{k \in K_i'} \delta_{ik} = 1.$$

Для пунктов K_i'', где сохранение действующих или строительство новых предприятий не является обязательным,

$$\sum_{k \subset K_i''} \delta_{ik} \leqslant 1.$$

Условие целочисленности:

$$\delta_{ik} = \delta_{ik}^2.$$

Рассмотрим модель с учетом размещения пунктов потребления. Пусть x_{ij} — объем перевозки однородного продукта из i-го пункта производства в j-й пункт потребления ($i = \overline{1,m}$; $j = \overline{1,n}$); c_{ij} — удельные затраты на такую перевозку. Обозначим через a_i максимально допустимую мощность i-го предприятия, через b_j — объем потребления продукции в

j-м пункте. Тогда задача развития и размещения производства сведется к минимизации затрат, связанных с производством и транспортировкой продукции:

$$\min Z = \sum_{i=1}^{m} \sum_{k=1}^{k_i} \alpha_{ik} \rho_{ik} \delta_{ik} + \sum_{i=1}^{m} \sum_{j=1}^{k_i} c_{ij} x_{ij}.$$

При этом спрос потребителей должен быть удовлетворен:

$$\sum_{i=1}^{m} x_{ij} \geqslant b_j \quad (j = \overline{1, n}),$$

а возможности поставщиков — ограничены:

$$\sum_{j=1}^{n} x_{ij} \leqslant a_i.$$

Если невозможно подобрать мощности так, чтобы потребность удовлетворялась при их полной загрузке, придется, учитывая дискретность мощностей, потребовать выполнения условия

$$\sum_{j=1}^{n} x_{ij} \leqslant \sum_{k=1}^{k_i} \rho_{ik} \delta_{ik} \quad (i = \overline{1, m}).$$

Условия целочисленности:

$$\sum_{k=1}^{k_i} \delta_{ik} \leqslant 1, \quad \delta_{ik} = \delta_{ik}^2 \qquad (i = \overline{1, m};\ k = \overline{1, k_i}),$$

условия неотрицательности:

$$x_{ij} \geqslant 0 \quad (i = \overline{1, m};\ j = \overline{1, n}).$$

Краткая классификация математических моделей дискретного программирования. Как следует из рассмотренных выше содержательных примеров, в экономической практике существуют задачи, которые формально к целочисленным не относятся. Требование целочисленности в них в явном виде не налагается. Но при целочисленных значениях некоторых исходных данных они обладают целочисленным оптимальным планом. Отнесем их к первому типу. Таким свойством обладают транспортная задача и ее модификации. Это свойство транспортной задачи легко объясняется, если вспомнить, что численные методы ее решения требуют применения к a_i и b_j лишь действий сложения и вычитания. Ко второму типу относятся задачи с неделимостями, например задача о рюкзаке, к третьему типу — модели с булевыми переменными. Примерами таких задач могут быть задача о назначениях, вариантные задачи размещения производительных сил и др. Сюда же относятся задачи типа задачи коммивояжера. Однако в силу ярко выраженных комбинаторных свойств последних их относят к четвертому типу — моделям задач комбинаторного типа. Из комбинаторных задач иногда выделяют класс задач о покрытии. Они касаются нахождения минимального подмножества множества ребер данного графа, содержащего все вершины графа. Такие задачи получили широкое применение при проектировании логических и вычислительных устройств.

В свою очередь каждый тип моделей подразделяется на линейные и нелинейные, статические или динамические, детерминированные или стохастические и т. д. Наиболее изучен класс задач целочисленного программирования детерминированного типа, в частности детерминированная задача дискретного линейного программирования. В общем виде такая задача имеет модель

$$\max(\min) Z = \sum_{j=1}^{n} c_{ij} x_j$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1, m}),$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}),$$

$$x_j \text{ — целые} \quad (j = J),$$

где J — некоторое подмножество множества индексов $j \in \in \{1, 2, \ldots, n\} \equiv J$. Если $J = \mathbf{N}$, то задачу называют *полностью целочисленной*, если же $J \subset \mathbf{N}$, — *частично целочисленной*.

Суть методов дискретной оптимизации. Во всех случаях решение задачи, казалось бы, может быть найдено обычными методами с отброшенными условиями целочисленности и последующими округлениями нецелых переменных в ответе. Однако такое округление может привести к решению, далекому от оптимального. Рассмотрим, например, геометрическую интерпретацию задачи дискретного линейного программирования (рис. 6.3).

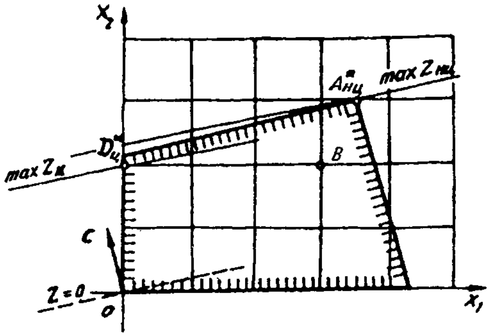


Рис. 6.3

Оптимальным решением нецелочисленной задачи служит точка $A^*_{нц}$. Как следует из рис. 6.3, точка B ближе всего к точке $A^*_{нц}$ в смысле округления. Но целой точкой, находящейся ближе всего к разрешающей линии уровня $\max Z_{нц}$, является точка $D^*_ц$. Таким образом, попытка решить задачу с отброшенным условием целочисленности и последующим округлением полученного оптимального плана до ближайших целых значений не всегда состоятельна. С практической точки зрения подобный подход допустим в тех случаях, когда значения переменных, образующих оптимальное решение исходной задачи, достаточно велики и погрешностями округления можно пренебречь.

В первом приближении методы целочисленной оптимизации можно разделить на две основные группы: точные и приближенные. К точным относятся методы отсечения и комбинаторные (метод ветвей и границ). Это универсальные методы дискретной оптимизации. Кроме универсальных, имеется много специальных точных методов, учитывающих специфику задачи. Однако точные методы имеют слабую сходимость. Многие экспериментальные и прикладные задачи не удалось решить точными методами за десятки и сотни тысяч итераций, хотя их конечность теоретически доказана. Трудности машинной реализации точных методов привели к появлению различного рода приближенных методов, построенных на использовании особенностей конкретной задачи. Среди приближенных методов наметились два направления: 1) разработка детерминированных эвристических алгоритмов, учитывающих специфику задачи; 2) использование случайного поиска в сочетании с локальной оптимизацией.

Общая идея решения задачи дискретного программирования методами отсечения состоит в следующем. Исходная задача решается сначала без учета ограничений целочисленности. Если полученный оптимальный план удовлетворяет условиям целочисленности, то задача решена. В противном случае к ограничениям исходной задачи добавляется новое, обладающее следующими свойствами: 1) полученный нецелочисленный план нарушает это ограничение; 2) любой целочисленный допустимый план исходной задачи заведомо удовлетворяет и новому ограничению. Затем задача решается с учетом нового ограничения. В случае необходимости добавляется еще одно ограничение и т. д. Геометрически каждому новому ограничению соответствует поверхность, которая отсекает от области допустимых решений некоторую его часть с оптимальной точкой с нецелыми координатами, но не затрагивает ни одной из целочисленных точек этого многогранника.

На основе этой идеи американский математик Р. Гомори предложил ряд сходящихся алгоритмов решения задач дискретного линейного программирования. Ему удалось обосновать правила построения дополнительных ограничений и доказать конечность алгоритмов.

Для решения задач дискретного (особенно нелинейного) программирования получили широкое распространение комбинаторные методы направленного частичного перебора допустимых планов. Из них наиболее универсален *метод ветвей и границ*, для выявления сути которого воспользуемся известной задачей "математической смекалки" об отыскании фальшивой монеты.

Пусть в мешке с монетами одинакового достоинства имеется одна фальшивая, отличающаяся бо́льшим весом, которую нужно отыскать посредством взвешивания на рычажных весах без гирь. Поступим так. Разделим содержимое мешка на две равные по количеству монет части. В случае, если число монет n нечетное, разделим $n-1$ монет на две равные части. Положим на чашки весов равные по количеству монет части. Если чашки весов уравновесятся, то отложенная монета фальшивая; в противном случае она находится в более тяжелой части, с которой поступим аналогичным образом, и т. д., пока не обнаружим фальшивую монету. Здесь деление мешка есть процесс ветвления множества на подмножества, т. е. разбиение области допустимых решений на непересекающиеся подмножества. Взвешивание каждой части соответствует оценке целевой функции на подмножествах (оценке их верхней или нижней границы). Если при этом удастся найти некоторый план, для которого верхняя (нижняя) оценка на множестве планов одного из подмножеств равна значению функции цели в этой точке и не меньше (не превосходит) оценок сверху (снизу) на всех подмножествах, то этот план оптимальный. Если такой план не обнаружен, то продолжаем процесс разбиения, начиная его с подмножества, имеющего самую высокую (низкую) оценку, и т. д. Поскольку множество всех планов задачи конечно, в конце концов оптимальный план будет найден.

6.2. МЕТОД ОТСЕЧЕНИЯ

Алгоритм метода Гомори для решения полностью целочисленной задачи линейного программирования. Рассмотрим полностью целочисленную задачу линейного программирования:

$$\max(\min) Z = \sum_{j=1}^{n} c_j x_j; \qquad (6.13)$$

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = \overline{1,m}), \tag{6.14}$$

$$x_j \geqslant 0 \quad (j = \overline{1,n}), \tag{6.15}$$

$$x_j \text{ — целые} \quad (j = \overline{1,n}). \tag{6.16}$$

Алгоритм Гомори состоит из следующих этапов.

1. Решается задача (6.13) — (6.15) с отброшенным условием целочисленности.

2. Полученное оптимальное решение (если оно существует) проверяется на целочисленность. Если условие целочисленности выполняется по всем переменным, то оптимальное решение задачи (6.13) — (6.16) совпадает с оптимальным решением задачи (6.13) — (6.15). Если это условие не выполняется хотя бы по одной переменной, то переходят к третьему этапу. Если задача (6.13) — (6.15) оказывается неразрешимой, то задача (6.13) — (6.16) тоже не имеет решения.

3. Строится дополнительное ограничение, отсекающее часть области, в которой содержится оптимальное решение задачи (6.13) — (6.15) и не содержится ни одного допустимого решения задачи (6.13) — (6.16).

4. Последний этап предусматривает возвращение к ЗЛП с отброшенным условием целочисленности, но с расширенной системой ограничений, в которую включено дополнительное ограничение, полученное на третьем шаге. К расширенной системе ограничений вновь применяется симплексная процедура. Если найденное таким образом решение будет опять нецелым, то формируется новое дополнительное ограничение и процесс вычислений повторяется.

Алгоритм Гомори позволяет за конечное число шагов прийти к оптимальному целочисленному решению, если оно существует.

С геометрической точки зрения каждому дополнительному линейному ограничению в n-мерном пространстве соответствует определенная гиперплоскость, отсекающая от многогранника решений некоторую его часть, включая и оптимальную на данном этапе нецелочисленную вершину. При этом все точки с целочисленными координатами, в том числе и искомая оптимальная, остаются в усеченном многограннике. Так как множество целых точек усеченного многогранника совпадает с множеством целых точек исходного многогранника, то понятно, что если оптимальное решение ЗЛП на усеченном многограннике удовлетворяет условию целочисленности, то оно является оптимальным целочисленным решением и исходной задачи. Через несколько операций отсечения искомая целочисленная точка оказывается сначала на границе, а затем становится оптимальной вершиной неоднократно усеченного многогранника решений.

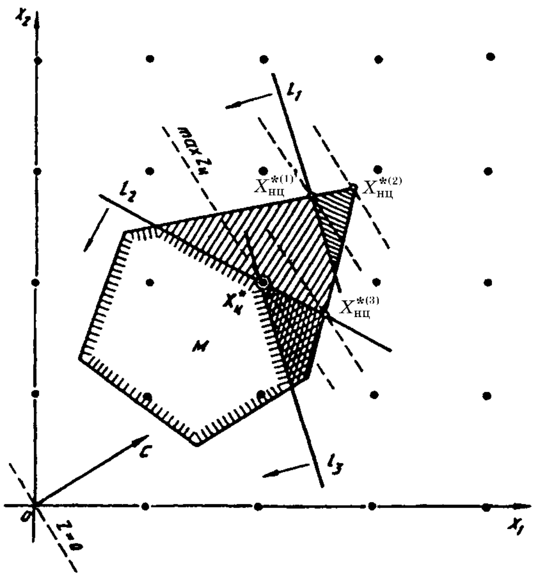


Рис. 6.4

Может оказаться, что многогранник решений не содержит ни одной целочисленной точки. В этом случае, сколько бы ни вводили дополнительных ограничений, целочисленного допустимого решения получить нельзя.

Для лучшего понимания сути вопроса обратимся к наглядной иллюстрации случая $n = 2$ (рис. 6.4). Ограничения задачи определяют на плоскости x_1Ox_2 некоторый многоугольник M, в вершине $\mathbf{x}^{*\,(1)}_{\text{нц}}$ которого, если не учитывать условия целочисленности, достигается максимум целевой функции. Внутри этого многоугольника имеется конечное множество точек, которым соответствуют решения с целочисленными значениями переменных (на рисунке они обозначены кружочками). На рисунке показаны три прямые l_1, l_2, l_3, соответствующие трем дополнительным линейным ограничениям. Каждая из прямых отсекает часть области допустимых решений. Так, после отсечения части области прямой оптимальной оказывается вершина $\mathbf{x}^{*}_{\text{нц}}{}^{(2)}$, затем $\mathbf{x}^{*}_{\text{нц}}{}^{(3)}$, и, наконец, оптимальной становится целочисленная вершина $\mathbf{x}^{*}_{\text{ц}}$. Основное в алгоритме — составление дополнительного ограничения, т. е. отсекающей гиперплоскости, которая называется *правильным отсечением*. Правильное отсечение должно удовлетворять следующим условиям: 1) быть линейным; 2) отсекать найденное оптимальное нецелочисленное решение задачи (6.13) — (6.15); 3) не отсекать ни одной из целочисленных точек задачи (6.13) — (6.16).

Формирование правильного отсечения. После каждой итерации система ограничений имеет вид

$$x_i = \beta_i - \sum_{x_j \in \{\text{СП}\}} \alpha_{ij} x_j, \quad x_i \in \{\text{БП}\}, \tag{6.17}$$

где $\{\text{БП}\}$ — множество базисных переменных.

Если выполняется условие оптимальности задачи, то находим оптимальное решение. Если все компоненты оптимального плана целочисленны, то задача решена. Предположим, что некоторые β_0 — нецелые. Пусть компонента i_0 — нецелая. Рассмотрим i_0-равенство системы (6.17), для которой выполняется условие оптимальности, т. е.

$$x_{i_0} = \beta_{i_0} - \sum_{x_j \in \{\text{СП}\}} \alpha_{i_0 j} x_j. \tag{6.18}$$

Напомним, что наибольшая целая часть числа a, его не превосходящая, обозначается $[a]$, а дробная положительная — $\{a\}$. Причем

$$a = [a] + \{a\}, \quad 0 \leqslant \{a\} < 1. \tag{6.19}$$

Например, пусть $a = 3,2$. Имеем $[3,2] = 3$, $\{3,2\} = 0,2$, $3,2 = 3 + 0,2$. Пусть $a = -4,3$. Имеем $[-4,3] = -5$, $\{-4,3\} = 0,7$, $-4,3 = -5 + 0,7$.

Перейдем к дальнейшему изучению уравнения (6.18). Найдем целую и дробную части его коэффициентов β_{i_0} и α_{i_0j}. Согласно равенству (6.19), имеем:

$$\beta_{i_0} = [\beta_{i_0}] + \{\beta_{i_0}\}, \tag{6.20}$$

$$\alpha_{i_0j} = [\alpha_{i_0j}] + \{\alpha_{i_0j}\}, \quad x_j \in \{\text{СП}\}.$$

Так как, по предположению, β_{i_0} нецелое, то $\{\beta_{i_0}\} > 0$. Кроме того, $\{\alpha_{i_0j}\} \geqslant 0$. Подставив выражение (6.20) в равенство (6.18), получим

$$x_{i_0} = \left([\beta_{i_0}] - \sum_{x_j \in \{\text{СП}\}} [\alpha_{i_0j}]x_j\right) + \left(\{\beta_{i_0}\} - \sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0j}\}x_j\right). \tag{6.21}$$

Так как первое слагаемое равенства (6.21) есть целое число, то, для того чтобы x_{i_0} было целым, необходимо, чтобы второе слагаемое также было целым, т. е. величина

$$L_{i_0} = \{\beta_{i_0}\} - \sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0j}\}x_j$$

должна быть целым числом. Так как x_{i_0} — координата допустимого целочисленного решения задачи (6.13) — (6.16), то L_{i_0} — всегда целое число. Покажем, что $L_{i_0} \leqslant 0$. В самом деле, величина

$$\sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0j}\}x_j$$

не может быть отрицательной. Из условия (6.19) следует $0 \leqslant \{\beta_{i_0}\} < 1$. Так как L_{i_0} — целое, то из предположения, что $L_{i_0} > 0$, должно следовать $\{\beta_{i_0}\} > 1$, что противоречит определению дробной части числа.

Итак, доказано, что любое допустимое решение задачи (6.13) — (6.16) должно удовлетворять неравенству

$$\{\beta_i\} - \sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0j}\}x_j \leqslant 0. \tag{6.22}$$

Теорема 6.1. *Неравенство (6.22) определяет правильное отсечение Гомори, т. е.: 1) является линейным; 2) отсекает найденное оптимальное нецелочисленное решение задачи*

(6.13) — (6.15); 3) не отсекает ни одного целочисленного плана задачи (6.13) — (6.16).

Доказательство. 1. Линейность соотношения (6.22) очевидна.

2. Пусть $\mathbf{x}^*_{\text{нц}}$ — оптимальный нецелочисленный план задачи (6.13) — (6.15), причем, например, координата x_{i_0} есть число нецелое. Покажем, что это решение не удовлетворяет условию (6.22). Поскольку $\mathbf{x}^*_{\text{нц}}$ оптимален, то $x^*_j = 0$, $x^*_j \in \{\text{СП}\}$. Поэтому

$$\sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0 j}\} x_j = 0. \tag{6.23}$$

Учитывая равенство (6.23), из условия (6.22) имеем $\{\beta_{i_0}\} \leqslant 0$, что противоречит определению дробной части числа. Итак, оптимальное решение $\mathbf{x}^*_{\text{нц}}$ задачи (6.13) — (6.15) не удовлетворяет условию (6.22).

3. Требование 3 по существу доказано при формировании условия (6.22). В самом деле, предположим, что существует целая точка $\mathbf{x}^*_{\text{ц}}$ задачи (6.13) — (6.16), которая не удовлетворяет условию (6.22), т. е.

$$\{\beta_{i_0}\} - \sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0 j}\} x^0_j \geqslant 0. \tag{6.24}$$

Так как $\sum\limits_{x_j \in \{\text{СП}\}} \{\alpha_{i_0 j}\} x_j \geqslant 0$ и $0 \leqslant \{\beta_{i_0}\} < 1$, то, учитывая неравенство (6.24), получаем

$$0 \leqslant \{\beta_{i_0}\} - \sum_{x_j \in \{\text{СП}\}} \{\alpha_{i_0 j}\} x^0_j < 1,$$

а это противоречит тому, что L_{i_0} для всех планов задачи (6.13) — (6.16) есть число целое. Соотношение (6.22) определяет правильное отсечение Гомори. $\triangle$

Если после очередной итерации окажется, что в оптимальном плане задачи (6.13) — (6.15) имеется несколько нецелых координат, то для построения отсекающей гиперплоскости целесообразно выбрать строку, содержащую свободный член с наибольшей дробной частью.

Признаком отсутствия целочисленного решения служит появление в симплексной таблице хотя бы одной строки с дробным свободным членом и целыми остальными коэффициентами, так как в этом случае соответствующее уравнение не имеет решения в целых числах.

Пример 6.4. Решить задачу

$$\max Z = x_1 + x_2;$$

$$\left.\begin{aligned} -x_1 + 3x_2 &\leqslant 3, \\ 3x_1 - x_2 &\leqslant 3, \end{aligned}\right\}$$

$$x_1, x_2 \geqslant 0 \text{ — целые.}$$

Решение. В результате решения задачи без учета условия целочисленности переменных получаем оптимальный план $\mathbf{x}^{*(1)}_{\text{нц}} = (3/2; 3/2; 0; 0)$, содержащийся в табл. 6.3.

Таблица 6.3

БП	$c_Б$	A_0	x_1	x_2	x_3	x_4
			1	1	0	0
x_2	1	3/2	0	1	3/8	1/8
x_1	1	3/2	1	0	1/8	3/8
$z_j - c_j$		3	0	0	1/2	1/2

Таблица 6.4

БП	$c_Б$	A_0	x_1	x_2	x_3	x_4	x_5
			1	1	0	0	0
x_2	1	3/2	0	1	3/8	1/8	0
x_1	1	3/2	1	0	1/8	3/8	0
—	—	4	0	0	1	$\boxed{3}$	-1
$z_j - c_j$		3	0	0	1/2	1/2	0
x_2	1	4/3	0	1	1/3	0	1/24
x_1	1	1	1	0	0	0	1/8
x_4	0	4/3	0	0	1/3	1	-1/3
$z_j - c_j$		7/3	0	0	1/3	0	1/6

В полученном плане две компоненты нецелые. Поскольку они равны по величине, сформируем правильное отсечение (6.22), например по x_1-строке: $\{3/2\} - \{1/8\}x_3 - \{3/8\}x_4 \leqslant 0$ или $x_3 + 3x_4 \geqslant 4$. Эквивалентное уравнение имеет вид $x_3 + 3x_4 - x_5 = 4$. Дополняя табл. 6.3 строкой для этого уравнения и столбцом для новой переменной x_5, получаем расширенную задачу, записанную в табл. 6.4. В этой же таблице в результате симплексного преобразования найден оптимальный нецелочисленный план $\mathbf{x}^{*}_{\text{нц}}{}^{(2)} = (1; 4/3; 0; 4/3; 0)$ расширенной задачи. Продолжаем процедуру отсечения.

Новое правильное отсечение (6.22) сформируем, например, по x_2-строке: $\{4/3\} - \{1/3\}x_3 - \{1/24\}x_5 \leqslant 0$ или $8x_3 + x_5 \geqslant 8$. Эквивалентное этому неравенству уравнение принимает вид $8x_3 + x_5 - x_6 = 8$.

Вновь расширяем задачу, добавляя строку для уравнения $8x_3 + x_5 - -x_6 = 8$ и столбец для переменной x_6. В результате табл. 6.4 преобразовывается в табл. 6.5.

Таблица 6.5

БП	$c_Б$	$\mathbf{A}_0$	x_1	x_2	x_3	x_4	x_5	x_5
			1	1	0	0	0	0
x_2	1	4/3	0	1	1/3	0	1/24	0
x_1	1	1	1	0	0	0	1/8	0
x_4	0	4/3	0	0	1/3	1	−1/3	0
—	—	8	0	0	$\boxed{8}$	0	1	−1
$z_j - c_j$		7/3	0	0	1/3	0	1/6	0
x_2	1	1						
x_1	1	1						
x_4	0	1						
x_3	0	1						
$z_j - c_j$		2	0	0	0	0	1/8	1/24

После очередной итерации получен оптимальный план $\mathbf{x}^{*}_{\text{ц}}{}^{(3)} = (1; 1; 1; 1; 0; 0)$ новой расширенной задачи. Все компоненты этого плана — целые числа. Максимальное значение целевой функции выражается целым числом: $\max Z_{\text{ц}} = 2$, а посему решение задачи закончено. Итак, для исходной задачи получили $\mathbf{x}^* = (1; 1)$, $Z^* = 2$.

Геометрическая иллюстрация решения задачи приведена на рис. 6.5. Первоначальная область допустимых решений ограничена осями координат и прямыми $-x_1 + 3x_2 = 3$ и $3x_1 - x_2 = 3$. Оптимальное нецелочисленное решение находится в вершине $\mathbf{x}^{*(1)}_{\text{нц}} = (3/2; 3/2)$.

Первому дополнительному ограничению $x_3 + 3x_4 \geqslant 4$ соответствует прямая l_1, уравнение $x_3 + 3x_4 = 4$ которой с учетом равенств $x_3 = = 3 + x_1 - 3x_2$ и $x_4 = 3 - 3x_1 + x_2$ приводится к виду $x_1 = 1$. В усеченном многоугольнике оптимум достигается в вершине $\mathbf{x}^{*(1)}_{\text{нц}} = (1; 4/3)$. Второму дополнительному ограничению $8x_3 + x_5 \geqslant 8$ соответствует прямая

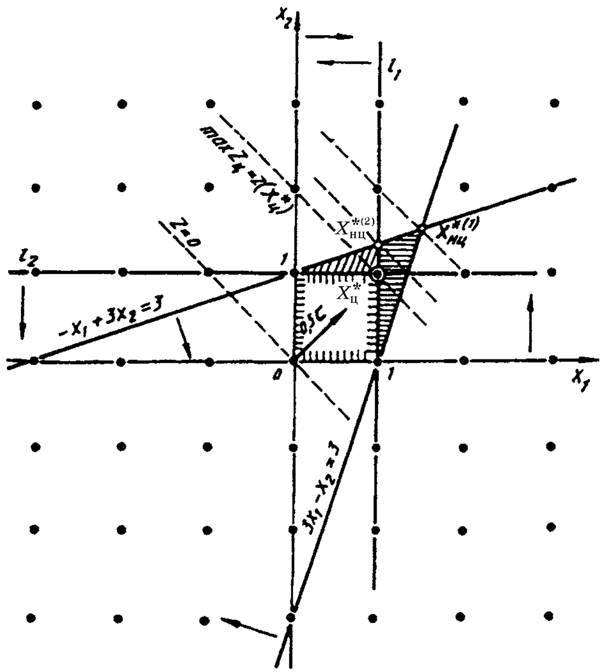


Рис. 6.5

l_2 с уравнением $x_2 = 1$. Дважды усеченная область решений превратилась в квадрат. Максимума в этом квадрате целевая функция достигает в целочисленной вершине $\mathbf{x}^*_{\text{ц}} = (1; 1)$. При этом $\max Z_{\text{ц}} = 2$.

6.3. МЕТОД ВЕТВЕЙ И ГРАНИЦ

Суть метода ветвей и границ. Метод ветвей и границ относится к группе комбинаторных методов. Комбинаторные методы исходят из конечности числа допустимых планов задачи и заменяют полный перебор всех планов их частичным направленным перебором. Комбинаторные методы в значительно меньшей степени подвержены в процессе вычислений влиянию ошибок округления, поэтому являются более предпочтительными по сравнению с методами отсечения. Метод ветвей и границ — один из наиболее эффективных методов решения задач комбинаторного типа.

Перейдем к изложению идеи метода ветвей и границ. Для этого рассмотрим общую задачу дискретного программирования

$$\max Z = f(\mathbf{x}), \tag{6.25}$$

$$\mathbf{x} \in \Omega, \tag{6.26}$$

где Ω — конечное множество допустимых планов.

1. Находим верхнюю границу (оценку) функции $f(\mathbf{x})$, $\mathbf{x} \in \Omega$, т. е. такое число $\varphi_0(\Omega)$, что для любых $\mathbf{x} \in \Omega$

$$f(\mathbf{x}) \leqslant \varphi_0(\Omega).$$

Если при этом удается найти такой план $\mathbf{x}_0$ задачи (6.25) — (6.26), для которого выполняется равенство

$$f(\mathbf{x}_0) = \varphi_0(\Omega),$$

то $\mathbf{x}_0$ — оптимальный план задачи (6.25) — (6.26).

2. Если оптимальный план не найден, то некоторым способом разбиваем множество Ω на конечное число непересекающихся подмножеств Ω_r^1:

$$\Omega = \bigcup_{r=1}^{r_1} \Omega_r^1, \quad \bigcap_{r=1}^{r_1} \Omega_r^1 = \varnothing$$

и находим для каждого их этих подмножеств верхнюю границу $\varphi_1(\Omega_r^1)$ $(r = \overline{1, r_1})$. Если при этом удастся найти такой план $\mathbf{x}_r \in \Omega_r^1$ $(1 \leqslant r \leqslant r_1)$, что выполняется соотношение

$$f(\mathbf{x}_l^1) = \varphi(\Omega_l^1) \geqslant \varphi(\Omega_r) \quad (r = \overline{1, r_1}),$$

то $\mathbf{x}^1$ — оптимальный план задачи (6.25) — (6.26). Если же такой план не найден, то выбираем подмножество Ω_r^1 с наибольшей верхней границей (перспективное подмножество) и разбиваем его на несколько непересекающихся подмножеств Ω_s^2 $(s = \overline{1, s_1})$. Для каждого нового подмножества находим верхнюю границу $\varphi(\Omega_s^2)$. Если будет найден такой план $\mathbf{x}_k^2$, что

$$f(\mathbf{x}_k^2) = \varphi(\Omega_k^2) \geqslant \varphi(\Omega_s^2),$$

то $\mathbf{x}_k^2$ — оптимальный план задачи. Если оптимальный план не найден, то дальнейшему ветвлению подвергаем подмножество с наибольшей верхней границей, и т. д. Процесс продолжается до получения оптимального плана. Способы ветвления и нахождения верхних границ выбираются для каждой конкретной задачи дискретного программирования. Процесс сопровождается построением дерева ветвления (рис. 6.6).

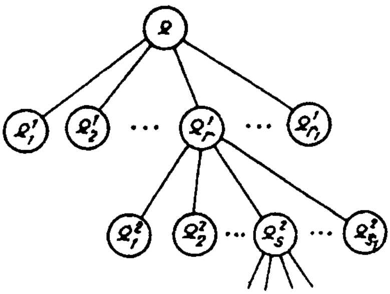


Рис. 6.6

Задача целочисленного (частично целочисленного) линейного программирования (6.13) — (6.16). В качестве верхней границы на множестве планов рассматривают значение целевой функции на оптимальном плане соответствующей ЗЛП (6.13) — (6.15). Пусть, далее, x_r — целочисленная переменная, значение x_r^* которой в оптимальном решении задачи (6.13) — (6.15) является дробным.

Интервал $[x_r^*] < x_r < [x_r^*] + 1$ не содержит допустимых целочисленных компонент решения. Поэтому допустимое целое значение x_r должно удовлетворять одному из неравенств: $x_r \leqslant [x_r^*]$ или $x_r \geqslant [x_r^*] + 1$. Введение этих условий в задачу с отброшенным условием целочисленности порождает две не связанные между собой задачи. Говорят, что исходная задача разветвляется на две подзадачи. Затем каждая подзадача решается как ЗЛП с целевой функцией исходной задачи

$$\max Z = \sum_{j=1}^{n} c_j x_j; \qquad \Bigg| \qquad \max Z = \sum_{j=1}^{n} c_j x_j;$$

$$\mathbf{x} \in \Omega,\ x_r \leqslant [x_r^*], \qquad \Bigg| \qquad \mathbf{x} \in \Omega,\ x_r \geqslant [x_r^*] + 1.$$

Если полученное оптимальное решение оказывается допустимым для целочисленной задачи, то его следует зафиксировать как наилучшее. При этом нет необходимости продолжать ветвление подзадачи, поскольку улучшить полученное решение не удастся. В противном случае подзадача должна быть разбита на две подзадачи и т. д. Как только полученное допустимое целочисленное решение одной из подзадач оказывается лучше имеющегося, оно фиксируется вместо зафиксированного ранее. Процесс ветвления продолжается до тех пор, пока каждая подзадача не приведет к целочисленному решению или пока не будет установлена невозможность улучшения имеющегося решения.

Пример 6.5. Решить задачу об оптимизации плана поставок крупногабаритной продукции:

$$\max Z = 2x_1 + 3x_2;$$

$$\left.\begin{aligned} 5x_1 + 7x_2 &\leqslant 35,\\ 4x_1 + 9x_2 &\leqslant 36, \end{aligned}\right\}$$

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_1, x_2 \text{ — целые.}$$

Решение. На рис. 6.7 изображена совокупность порожденных подзадач в виде дерева ветвления с начальной вершиной *0*, в которой исходная задача решается как ЗЛП с отброшенным условием целочисленности. Оптимальное решение достигается в точке $x^*_{\text{нц}} = \left(3\frac{12}{17}; 2\frac{6}{17}\right)$.

Так как обе переменные принимают нецелые значения, то любая из них может быть выбрана в качестве переменной, продолжающей процесс ветвления.

Выбор, например, переменной x_2 порождает две подзадачи, связанные с условием $x_2 \leqslant [x_2^*]$ или $x_2 \geqslant [x_2^*] + 1$. Так как $[x_2^*] = [40/17] = 2$, имеем две подзадачи *1.1* и *1.2* (так же будем обозначать вершины дерева ветвления):

1.1

$$\max Z = 2x_1 + 3x_2;$$

$$\left.\begin{aligned} 5x_1 + 7x_2 &\leqslant 35,\\ 4x_1 + 9x_2 &\leqslant 36,\\ x_2 &\leqslant 2, \end{aligned}\right\}$$

$$x_1, x_2 \geqslant 0 \text{ — целые,}$$

1.2

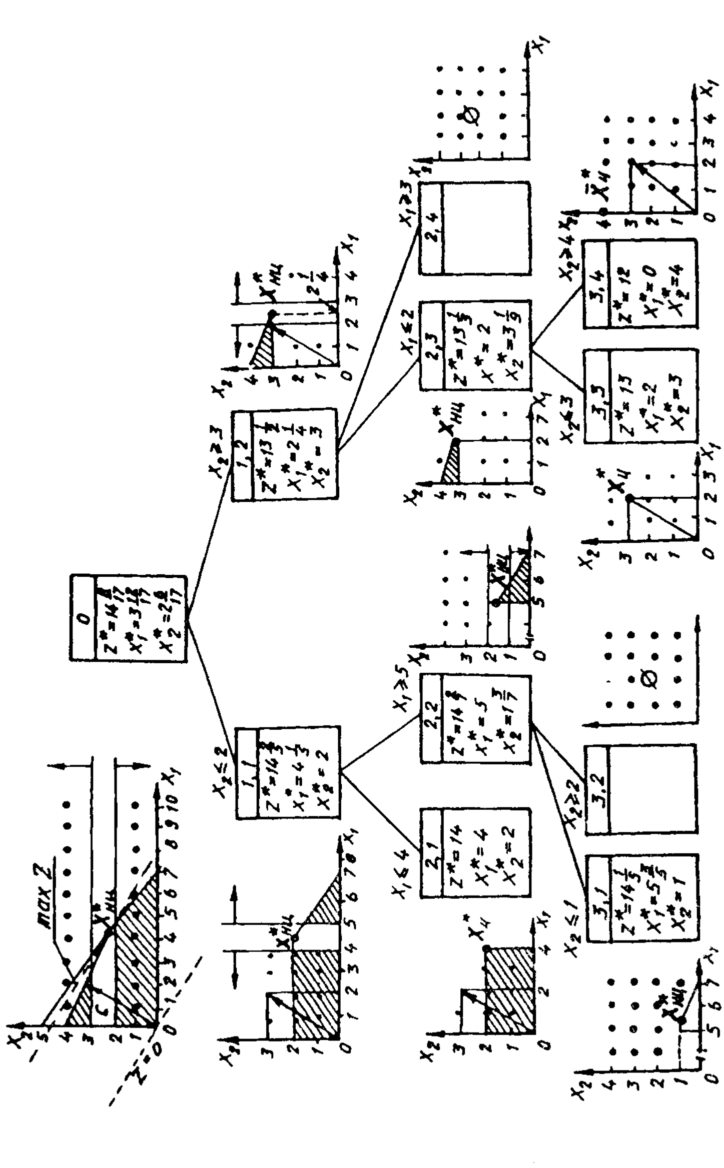
$$\max Z = 2x_1 + 3x_2;$$

$$\left.\begin{aligned} 5x_1 + 7x_2 &\leqslant 35,\\ 4x_1 + 9x_2 &\leqslant 36,\\ x_2 &\geqslant 3, \end{aligned}\right\}$$

$$x_1, x_2 \geqslant 0 \text{ — целые.}$$

Порожденные подзадачи содержат все допустимые целочисленные решения исходной задачи, т. е. исходное множество допустимых целочисленных решений остается неизменным в процессе ветвления. Их решения представлены на рис. 6.7.

На следующем шаге осуществляется выбор одной из подзадач *1.1* или *1.2* для решения и при необходимости для дальнейшего ветвления.



max Z
0
1.1
1.2
2.1
2.2
2.3
2.4
3.1
3.2
3.3
3.4
$x_2 \leq 2$
$x_2 \geq 3$
$x_1 \leq 4$
$x_1 \geq 5$
$x_1 \leq 2$
$x_1 \geq 3$
$x_2 \leq 1$
$x_2 \geq 2$
$x_2 \leq 3$
$x_2 \geq 4$
$Z^* = 14$
$x_1^* = 4$
$x_2^* = 2$
$Z^* = 13$
$x_1^* = 2$
$x_2^* = 3$
$Z^* = 12$
$x_1^* = 0$
$x_2^* = 4$

Рис. 6.7

Отметим, что не существует точных способов реализации указанного выбора. Причем выбор различных альтернатив приводит к разным последовательностям подзадач, а следовательно, к различным количествам итераций, обеспечивающих получение оптимального целочисленного решения. Проиллюстрируем этот факт, используя рис. 6.7. Предположим, что в первую очередь рассматривается вершина *1.1*. Оптимальное решение этой задачи достигается в точке $x^*_{\text{нц}} = \left(4\frac{1}{5}; 2\right)$. Так как значение $x_1^* = 4\frac{1}{5}$ остается нецелым, задача *1.1* порождает подзадачи *2.1* и *2.2* с дополнительными ограничениями соответственно $x_1 \leqslant [x_1^*]$, т.е. $x_1 \leqslant 4$ и $x_1 \geqslant [x_1^*] + 1$, т.е. $x_1 \geqslant 5$:

2.1	*2.2*
$\max Z = 2x_1 + 3x_2;$	$\max Z = 2x_1 + 3x_2;$
$\left.\begin{array}{c} 5x_1 + 7x_2 \leqslant 35, \\ 4x_1 + 9x_2 \leqslant 36, \\ x_1 \leqslant 4,\ x_2 \leqslant 2, \end{array}\right\}$	$\left.\begin{array}{c} 5x_1 + 7x_2 \leqslant 35, \\ 4x_1 + 9x_2 \leqslant 36, \\ x_1 \geqslant 5,\ x_2 \leqslant 2, \end{array}\right\}$
$x_1, x_2 \geqslant 0$ — целые,	$x_1, x_2 \geqslant 0$ — целые.

Решения этих подзадач приведены на рис. 6.7.

Наличие у подзадачи *2.1* целочисленного решения не означает, что найдено оптимальное целочисленное решение исходной задачи *0*, потому что еще не решены подзадачи *1.2* и *2.2*, которые могут дать лучшее решение, чем *2.1*. Целочисленное решение подзадачи *2.1* определяет нижнюю границу $Z = 14$ значений целевой функции. Нет необходимости рассматривать те последующие подзадачи, для которых оптимальные значения Z меньше указанной нижней границы.

Обратимся к подзадаче *1.2*. Для нее $Z^* = 13,5$, что не превышает значения $Z = 14$, поэтому поиск вдоль ветви $x_2 \geqslant 3$ следует прекратить.

Исследуем вершину *2.2*, которой соответствует $Z^* = 14\frac{2}{7}$. Несмотря на то что полученное значение превышает нижнюю границу $Z = 14$, дальнейшее ветвление осуществлять нецелесообразно, поскольку $14\frac{2}{7} - 14 < 1$ и все коэффициенты целевой функции целые. Продолжение ветвления подзадачи *3.1* показывает, что $Z^* = 14\frac{1}{5}$, что меньше $Z^* = 14\frac{2}{7}$ в вершине *2.2*. Вторая ветвь *3.2* дает пустое множество допустимых решений. Итак, нашли, наконец, оптимальное решение задачи: $\mathbf{x}^*_{\text{ц}} = (4; 2)$, $z(\mathbf{x}^*_{\text{ц}}) = 14$. К нему приведена цепочка задач $0 \to 1.1 \to 1.2 \to 2.1 \to 2.2 \to 3.1 \to 3.2$. Отметим, что при движении по этой цепочке в вершинах *1.2* и *2.2* мы не имеем сведений о качестве полученного ранее решения, выраженного значением целевой функции Z, до тех пор, пока не решили подзадачи, соответствующие указанным вершинам. По этой причине нельзя предугадать, является ли выбор ветви *1.2* ($x_2 \geqslant 3$), осуществленный в вершине *0*, более выгодным, чем выбор ветви *1.1* ($x_2 \leqslant 2$), или наоборот. Для иллюстрации этого факта предположим, что сначала исследуется вершина *1.2* ($x_2 \geqslant 3$). Так как значение x_1^* не является целым, то исходящие из вершины *1.2* ($x_2 \leqslant 3$) ветви $x_1 \leqslant 2$ и $x_1 \geqslant 3$ порождают подзадачи *2.3* ($x_1 \leqslant 2$) и *2.4* ($x_1 \geqslant 3$). Пусть на следующем шаге решается подзадача *2.3* ($x_1 \leqslant 2$). Получаем $\mathbf{x}^*_{\text{нц}} = \left(2; 3\frac{1}{9}\right)$,

$Z^* = 13\frac{1}{3}$. Возникают две новые ветви ($x_2 \leqslant 3$ и $x_2 \geqslant 4$) с подзадачами *3.3* и *3.4*, имеющими следующие оптимальные целочисленные решения: $\mathbf{x}^*_{\text{ц}} = (2; 3)$, $Z^* = 13$ и $\mathbf{x}^*_{\text{ц}} = (0; 4)$, $Z^* = 12$. Они явно хуже, чем решение подзадачи *2.1*.

Рассмотренный пример наглядно показывает, что число итераций и рассматриваемых подзадач зависит от выбора вершины ветвления.

Решение задачи коммивояжера методом ветвей и границ (алгоритм Литтла). Модель задачи — (6.7) — (6.10). Если считать города вершинами графа, а коммуникации (i, j) — его дугами, то требование нахождения минимального пути, проходящего один и только один раз через каждый город, и возвращения в него можно рассматривать как нахождение на графе контура минимальной длины. Замкнутый контур, проходящий один и только один раз через каждую вершину графа, называется *гамильтоновым контуром.* Таким образом, задача коммивояжера состоит в нахождении на графе гамильтонова контура минимальной длины.

Рассмотрим *алгоритм Литтла* для нахождения минимального гамильтонова контура на графе с n вершинами. Если между вершинами i и j нет дуги, то ставится символ ∞. Этот же символ ставится на главной диагонали, что означает запрет на возвращение в вершину, через которую уже проходил контур. Основная идея метода состоит в том, что вначале строят нижнюю границу длин множества гамильтоновых контуров Ω^0. Затем множество контуров Ω^0 разбивается на два подмножества таким образом, чтобы первое подмножество Ω^1_{ij} состояло из гамильтоновых контуров, содержащих некоторую дугу (i, j), а другое подмножество $\Omega^1_{\overline{ij}}$ не содержало этой дуги. Для каждого из подмножеств определяются нижние границы по тому же правилу, что и для первоначального множества гамильтоновых контуров. Полученные нижние границы подмножеств Ω^1_{ij} и $\Omega^1_{\overline{ij}}$ оказываются не меньше нижней границы для всего множества гамильтоновых контуров, т. е.

$$\varphi(\Omega^0) \leqslant \varphi(\Omega^1_{ij}) \equiv \varphi^1_{ij}, \quad \varphi(\Omega^0) \leqslant \varphi(\Omega^1_{\overline{ij}}) \equiv \varphi^1_{\overline{ij}}.$$

Сравнивая нижние границы φ^1_{ij} и $\varphi^1_{\overline{ij}}$ подмножеств Ω^1_{ij} и $\Omega^1_{\overline{ij}}$, можно выделить среди них то, которое с большей вероятностью содержит гамильтонов контур минимальной длины. Затем одно из подмножеств Ω^1_{ij} или $\Omega^1_{\overline{ij}}$ по аналогичному правилу разбивается на два новых Ω^2_{ij} и $\Omega^2_{\overline{ij}}$. Для них снова отыскиваются нижние границы φ^2_{ij} и $\varphi^2_{\overline{ij}}$ и т. д. Процесс ветвления продолжается до тех пор, пока не отыщется единственный гамильтонов контур. Его называют *первым рекордом*. Затем просматривают оборванные ветви. Если их нижние границы больше длины первого рекорда, то задача решена, если же есть такие, для которых нижние границы меньше, чем длина первого рекорда, то подмножество с наименьшей нижней границей подвергается дальнейшему ветвлению, пока не убеждаются, что оно не содержит лучшего гамильтонова контура. Если же такой найдется, то анализ оборванных ветвей продолжается относительно нового значения длины контура. Его называют *вторым рекордом*. Процесс решения заканчивается, когда будут проанализированы все подмножества.

Для практической реализации идеи метода ветвей и границ применительно к задаче коммивояжера нужно указать прием определения нижних границ подмножеств и разбиения множества гамильтоновых контуров на подмножества (ветвление). Определение нижних границ базируется на следующем простом утверждении, которое сформулируем в виде теоремы.

Теорема 6.2. *Если к элементам любого ряда матрицы задачи коммивояжера (строке или столбцу) прибавить или вычесть из них некоторое число, то от этого оптимальность плана не изменится. Длина же любого гамильтонова контура изменится на данную величину.* Иначе: *если план* $\mathbf{x}^*$ *— оптимальный план задачи (6.7) — (6.10), то он также оптимален для задачи с функцией цели* Z' *и с матрицей*

$$C' = [c_{ij} \pm u_i \pm v_j] \quad (i, j = \overline{1, n}), \quad \mathbf{x} \in \Omega.$$

Доказательство. В самом деле,

$$Z' = \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} \pm u_i \pm v_j) x_{ij} =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \pm \sum_{i=1}^{n} u_i \sum_{j=1}^{n} x_{ij} \pm \sum_{j=1}^{n} v_j \sum_{i=1}^{n} x_{ij} =$$

$$= Z \pm \sum_{i=1}^{n} u_i \pm \sum_{j=1}^{n} v_j = Z + l,$$

где $l = \pm \sum\limits_{i=1}^{n} u_i \pm \sum\limits_{j=1}^{n} v_j$, т. е. $Z' = Z + l$. Очевидно, что Z и Z' достигают экстремального значения для одного и того же плана. $\triangle$

Опираясь на теорему 6.2, вычтем из каждой строки матрицы коммивояжера C число u_i, равное минимальному элементу этой строки. Перейдем к матрице C' с элементами $c'_{ij} = c_{ij} - \min\limits_{j} c_{ij}$. Вычтем затем из каждого j-го столбца матрицы C' число v_j, равное минимальному элементу этого столбца. Получим матрицу C'' с элементами $c''_{ij} = c_{ij} - \min\limits_{j} c_{ij} - \min\limits_{i} c'_{ij}$. Матрица C'' называется *приведенной* по строкам и столбцам. В приведенной матрице C'' в каждой строке и каждом столбце содержится хотя бы один нуль. Все элементы матрицы C'' неотрицательны: $c''_{ij} \geqslant 0$.

Величина $\sum\limits_{i=1}^{n} u_i + \sum\limits_{j=1}^{n} v_j = \varphi_0$ называется *константой приведения.*

Теорема 6.3. *Пусть дана задача коммивояжера с матрицей $C = [c_{ij}]$. Константа приведения этой матрицы φ_0 определяет нижнюю границу гамильтоновых контуров $\varphi(\Omega^0) = \varphi_0$.*

Доказательство. В самом деле, из теоремы 6.2 следует, что $Z'' = Z - \varphi_0$. Отсюда $Z = Z'' + \varphi_0$. Так как в приведенной матрице $c''_{ij} \geqslant 0$, то $Z'' \geqslant 0$. Отсюда $Z \geqslant \varphi_0$, т. е. φ_0 можно выбрать в качестве нижней границы всех гамильтоновых контуров $\varphi(\Omega^0) = \varphi_0$. $\triangle$

Рассмотрим вопрос о разбиении множества гамильтоновых контуров на подмножества. Это делается с помощью некоторой перспективной дуги (i_0, j_0). Первое подмножество $\Omega^1_{i_0 j_0}$ составляют гамильтоновы контуры, содержащие перспективную дугу (i_0, j_0), второе подмножество $\Omega^1_{\overline{i_0 j_0}}$ не содержит дуги (i_0, j_0).

Включение дуги (i_0, j_0) в контур приводит к автоматическому сокращению матрицы C на строку i_0 и столбец j_0. Так как гамильтонов контур не может одновременно содержать дуги (i_0, j_0) и (j_0, i_0), то, включив (i_0, j_0) в контур, следует (j_0, i_0) исключить, т. е. путь (j_0, i_0) "блокировать бесконечностью" $(c_{j_0 i_0} = \infty)$. Сократив размер матрицы, можно улучшить нижнюю границу подмножества гамильтоновых контуров.

Для выделения претендентов на включение в множество дуг, по которым производится ветвление, рассмотрим в приведенной матрице C'' все элементы, равные нулю $(c''_{ij} = 0)$. Найдем степени θ_{ij} нулевых элементов этой матрицы. Пусть $c''_{i_0 j_0} = 0$. Степень нулевого элемента равна сумме минимальных элементов в строке i_0 и столбце j_0 при блокировании перехода (i_0, j_0) бесконечностью, т. е.

$$\theta_{i_0 j_0} = \min_{j \neq j_0} c_{i_0 j} + \min_{i \neq i_0} c_{i j_0}.$$

Можно доказать (примем без доказательства), что с наибольшей вероятностью искомому гамильтонову контуру принадлежат дуги с максимальной степенью нуля. Пусть $\max\limits_{c''_{ij}=0} \theta_{ij} = \theta_{i_0 j_0}$. Дуга (i_0, j_0) включается в множество $\Omega^1_{i_0 j_0}$ и является запретной для множества $\Omega^1_{\overline{i_0 j_0}}$. Для получения матрицы контуров $\Omega^1_{i_0 j_0}$, включающих дугу (i_0, j_0), вычеркиваем в матрице C'' строку i_0 и столбец j_0, а чтобы не допустить образования негамильтонова контура, заменим симметричный относительно главной диагонали элемент (j_0, i_0) на ∞. Множество гамильтоновых контуров $\Omega^1_{\overline{i_0 j_0}}$, не включающих дугу (i_0, j_0), получаем путем замены элемента $c''_{i_0 j_0}$ в матрице C'' на ∞. Далее приводим матрицы подмножеств $\Omega^1_{i_0 j_0}$ и $\Omega^1_{\overline{i_0 j_0}}$ и определяем их нижние границы. Пусть $h^1_{i_0 j_0}$ и $h^1_{\overline{i_0 j_0}}$ — константы приведения матриц множеств $\Omega^1_{i_0 j_0}$ и $\Omega^1_{\overline{i_0 j_0}}$. Тогда:

$$\varphi^1_{i_0 j_0} = \varphi(\Omega^1_{i_0 j_0}) = \varphi(\Omega^0) + h^1_{i_0 j_0},$$
$$\varphi^1_{\overline{i_0 j_0}} = \varphi(\Omega^1_{\overline{i_0 j_0}}) = \varphi(\Omega^0) + h^1_{\overline{i_0 j_0}}.$$

Производим ветвление подмножества с меньшей нижней границей и т. д. Найдя таким образом хотя бы один гамильтонов контур (первый рекорд), сравним его длину с нижними границами оборванных ветвей. Процесс ветвления сопровождается построением дерева ветвления. Если длина контура не превышает нижних границ оборванных ветвей дерева, то задача решена. В противном случае ветвление продолжается с подмножества, имеющего наименьшую нижнюю границу.

Пример 6.6. Решить задачу определения оптимальной последовательности запуска пяти деталей в производство, если матрица потерь C от переналадок представлена в табл. 6.2, т. е. определить последовательность запуска деталей в производство, минимизирующую потери от переналадок.

Изобразим последовательность обработки партии деталей в виде графа, каждая вершина которого будет соответствовать определенной детали, а каждая дуга — операции переналадки станка с обработки одной детали на другую. Каждой дуге сопоставим число, характеризующее время, затрачиваемое на переналадку. В таких случаях проблема наиболее рациональной переналадки сведется к поиску на графе гамильтонова контура наименьшей продолжительности, т. е. по существу к решению задачи коммивояжера.

Решение. 1. Присоединим справа к матрице (табл. 6.2) столбец $u_i = \min\limits_j c_{ij}$ (табл. 6.6). Вычтем элементы u_i из соответствующих элементов матрицы C и получим матрицу C', приведенную по строкам (табл. 6.7):

$$C' = \left[c'_{ij}\right] = \left[c_{ij} - \min_j c_{ij}\right].$$

Таблица 6.6

$i \backslash j$	1	2	3	4	5	u_i
1	∞	35	45	20	11	11
2	9	∞	17	6	8	6
3	21	31	∞	2	11	2
4	30	15	40	∞	10	10
5	10	9	8	7	∞	7
					$\sum\limits_i u_i =$	36

Таблица 6.7

$i \backslash j$	1	2	3	4	5	
1	∞	24	34	9	0	
2	3	∞	11	0	2	
3	19	29	∞	0	9	
4	20	5	30	∞	0	
5	3	2	1	0	∞	
v_j	3	2	1	0	0	$\sum\limits_j v_j = 6$

2. Присоединим снизу к матрице (табл. 6.7) строку $v_j = \min\limits_j c'_{ij}$. Вычтя элементы v_j из соответствующих элементов столбцов матрицы C', получим матрицу $C'' = [c'_{ij} - \min c'_{ij}]$. Такая матрица называется приведенной по строкам и столбцам и обозначена $\widetilde{\Omega}^0$. В каждой строке и каждом столбце такой матрицы имеется хотя бы один нуль (табл. 6.8).

3. Найдем константу приведения:

$$\varphi_0 = \sum_i u_i + \sum_j v_j = 36 + 6 = 42.$$

Она определяет нижнюю границу множества всех гамильтоновых контуров, т. е.

$$\varphi_0 = \varphi(\Omega^0) \leqslant z(\mathbf{x}), \quad \mathbf{x} \in \Omega.$$

4. Найдем степени нулей для приведенной по строкам и столбцам матрицы $\widetilde{\Omega}^0$ (см. табл. 6.8). Для этого нуль в матрице $\widetilde{\Omega}^0$, степень которого отыскивается, мысленно заменяем на ∞ и находим сумму минимальных элементов строки и столбца, где он находится.

5. Выбираем дугу (i_0, j_0), для которой степень нулевого элемента наибольшая. В нашем случае она равна 10 и соответствует дуге (*5*, *3*). Таким образом, претендентом на включение в гамильтонов контур является дуга (*5*, *3*).

6. Разбиваем множество всех гамильтоновых контуров $\widetilde{\Omega}^0$ на два: Ω^1_{53} и $\Omega^1_{\overline{53}}$. Матрицу Ω^1_{53}, содержащую дугу (*5*, *3*), получаем из табл. 6.8 путем вычеркивания строки *5* и столбца *3*. Чтобы не допустить образования негамильтонова контура, заменяем элемент C''_{35} на ∞ (табл. 6.9).

Таблица 6.8

$\widetilde{\Omega}^0$

i \ j	*1*	*2*	*3*	*4*	*5*
1	∞	22	33	9	0^9
2	0^0	∞	10	0^0	2
3	16	27	∞	0^9	9
4	17	3	29	∞	0^3
5	0^0	0^3	0^{10}	0^0	∞

Таблица 6.9

$\widetilde{\Omega}^1_{53}$

i \ j	*1*	*2*	*4*	*5*
1	∞	22	9	0
2	0	∞	0	2
3	16	27	0	∞
4	17	3	∞	0
		3		

7. Выполняем дополнительное приведение матрицы контуров Ω^1_{53}. Находим $h^1_{53} = 3$ (табл. 6.10). Следовательно, нижняя граница множества $\widetilde{\Omega}^1_{53}$ равна $\varphi^1_{53} = 42 + 3 = 45$.

Таблица 6.10

$\widetilde{\Omega}^1_{53}$

i \ j	*1*	*2*	*4*	*5*
1	∞	19	9	0^9
2	0^{16}	∞	0^0	2
3	16	24	0^{16}	∞
4	17	0^{19}	∞	0^0

8. Матрицу гамильтоновых контуров $\Omega^1_{\overline{53}}$ получим из табл. 6.8 путем замены элемента c''_{53} на ∞ (табл. 6.11).

9. Находим дополнительную константу приведения для множества контуров $\Omega^1_{\overline{53}}$: $h^1_{\overline{53}} = 10$. Следовательно, нижняя граница этого множества $\varphi(\widetilde{\Omega}^1_{\overline{53}}) = 42 + 10 = 52$ (табл. 6.12).

Таблица 6.11

$\widetilde{\Omega}^1_{\overline{53}}$

i \ j	1	2	3	4	5
1	∞	22	33	9	0
2	0	∞	10	0	2
3	16	27	∞	0	9
4	17	3	29	∞	0
5	0	0	∞	0	∞
			10		

Таблица 6.12

$\widetilde{\Omega}^1_{\overline{53}}$

i \ j	1	2	3	4	5
1	∞	22	23	9	0
2	0	∞	0	0	2
3	16	27	∞	0	9
4	17	3	19	∞	0
5	0	0	∞	0	∞

10. Сравниваем нижние границы подмножеств $\widetilde{\Omega}^1_{53}$ и $\widetilde{\Omega}^1_{\overline{53}}$. Так как $\varphi^1_{53} = 45 < \varphi^1_{\overline{53}} = 52$, дальнейшему ветвлению подвергаем подмножество Ω^1_{53} (табл. 6.10). Процесс отыскания оптимального плана сопровождается построением дерева ветвления (рис. 6.8). На рис. 6.9 представлено ветвление по дуге (5, 3).

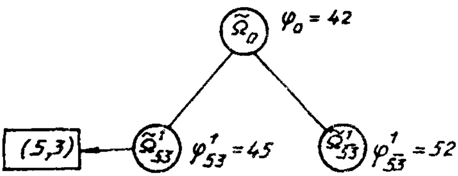


Рис. 6.8

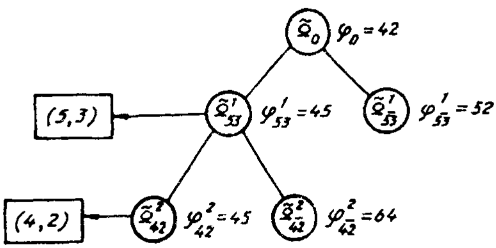


Рис. 6.9

Далее переходим к ветвлению подмножества $\tilde{\Omega}_{53}^1$. Находим степени нулей этой матрицы (см. табл. 6.10). Претендентом на включение в гамильтонов контур является дуга (4, 2) с наибольшей степенью нуля. Разбиваем множество $\tilde{\Omega}_{53}^1$ на два подмножества: Ω_{42}^2 и $\Omega_{\overline{42}}^2$ (табл. 6.13 и 6.14).

Таблица 6.13

$\tilde{\Omega}_{42}^2$

i \ j	1	4	5
1	∞	9	0^{11}
2	0^{18}	0	2
3	16	0^{25}	∞

Таблица 6.14

$\tilde{\Omega}_{\overline{42}}^2$

i \ j	1	2	4	5
1	∞	19	9	0
2	0	∞	0	2
3	16	24	0	∞
4	17	∞	∞	0
		19		

Определяем константы приведения для этих матриц: $h_{42}^2 = 0$, $h_{\overline{42}}^2 = 19$. Следовательно, $\varphi_{42}^2 = 45$, $\varphi_{\overline{42}}^2 = 45 + 19 = 64$. На рис. 6.10 представлено ветвление с использованием дуги (4, 2). Так как $\varphi_{42}^2 < \varphi_{\overline{42}}^2$,

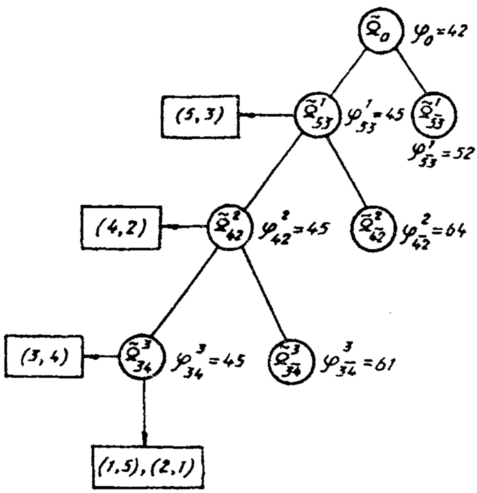


Рис. 6.10

ветвлению подлежит подмножество $\widetilde{\Omega}_{42}^2$ (см. табл. 6.13). Вычисляем степени его нулей. Претендентом на включение в гамильтонов контур является дуга (*3*, *4*). Разбиваем множество $\widetilde{\Omega}_{42}^2$ на подмножества Ω_{34}^3, $\Omega_{\overline{34}}^3$. Их матрицы представлены в табл. 6.15 и 6.16.

Таблица 6.15

$\widetilde{\Omega}_{34}^3$

$i \backslash j$	1	5
1	∞	0
2	0	2

Таблица 6.16

$\widetilde{\Omega}_{\overline{34}}^3$

$i \backslash j$	1	4	5
1	∞	9	0
2	0	0	2
3	16	∞	∞

Так как $\varphi_{34}^3 = 45 < \varphi_{\overline{34}}^3 = 61$, ветвлению нужно подвергнуть подмножество $\widetilde{\Omega}_{34}^3$. Но его матрица имеет размерность 2×2. Поэтому в гамильтонов контур следует включить дуги, соответствующие в матрице подмножества $\widetilde{\Omega}_{34}^3$ нулевым элементам, т. е. дуги (*1*, *5*) и (*2*, *1*). На рис. 6.10 представлено полное дерево ветвлений. Определим полученный гамильтонов контур: $\Gamma_1 = \langle (5,3),(4,2),(3,4),(1,5),(2,1)\rangle = \langle 5 \to 3 \to 4 \to 2 \to 1 \to 5\rangle$. Его длина

$$z(\Gamma_1) = c_{53} + c_{34} + c_{42} + c_{21} + c_{15} = 8 + 2 + 15 + 9 + 11 = 45.$$

Так как нижние границы оборванных ветвей больше длины первого рекорда $z(\Gamma^*) = 45$ (см. рис. 6.10), этот контур имеет наименьшую длину. На рис. 6.11 представлен замкнутый маршрут минимальной длины.

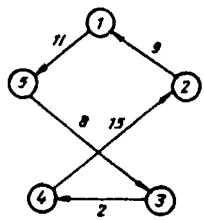


Рис. 6.11

Итак, партию из пяти деталей следует обрабатывать в следующем порядке: сначала первую, затем пятую, после нее третью, далее четвертую и, наконец, вторую.

Решение задачи коммивояжера на максимум. В случае, если каждая дуга графа связана с получением, например, прибыли, задачу коммивояжера нужно решать на максимум.

Задачу на максимум легко свести к алгоритму Литтла для решения задачи коммивояжера на минимум. Для этого найдем максимальный элемент в каждом столбце: $l_j = \max_i c_{ij}$. Построим матрицу $C' = [c'_{ij}] = [\max_i c_{ij} - c_{ij}]$. Имеем $Z' = \sum_{j=1}^{n} l_j - Z$. Отсюда следует, что Z' достигает минимума для того же плана, для которого Z достигает максимума. В случае решения задачи коммивояжера на максимум в исходной матрице по главной диагонали ставится символ $-\infty$. Этот же символ приписывается дугам, по которым нет перемещения.

7. ВЫПУКЛОЕ ПРОГРАММИРОВАНИЕ

7.1. МАТЕМАТИЧЕСКИЕ ОСНОВЫ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

Производная по направлению. Пусть дана функция $z = f(x, y)$, определенная в замкнутой области D плоскости xOy (рис. 7.1). Известно, что частные производные $\partial z/\partial x$ и $\partial z/\partial y$ функции $z = f(x, y)$ могут быть истолкованы как скорости изменения этой функции в точке $M_0(x, y)$ в направлении осей Ox и Oy.

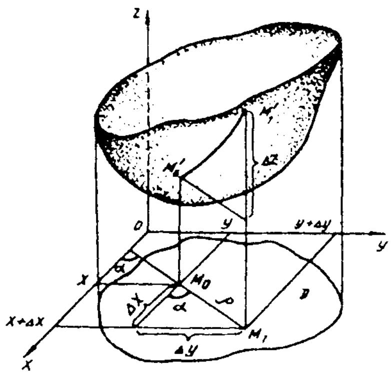


Рис. 7.1

Понятие производной можно обобщить, если рассматривать скорость изменения функции $z = f(x, y)$ в произвольном направлении, образующем некоторый угол с осью Ox. Пусть $M_0'(x, y, z)$ — точка на поверхности $z = f(x, y)$. Дадим аргументам x и y приращения Δx и Δy. Тогда функция z получит приращение Δz. Через M_1' обозначим точку на поверхности с координатами $x + \Delta x$, $y + \Delta y$, $z + \Delta z$; $M_0(x, y)$ и $M_1(x + \Delta x, y + \Delta y)$ — проекции точек M_0' и M_1' на плоскость xOy. Обозначим через α угол, который составляет прямая M_0M_1 с осью Ox $(0 \leqslant \alpha \leqslant 2\pi)$, а через ρ — длину отрезка M_0M_1.

Определение 7.1. *Производной $dz/d\rho$ функции $z = f(x, y)$ в точке $M_0(x, y)$ по направлению α называется предел отношения приращения Δz функции z, возникшего при перемещении точки M из положения M_0 вдоль луча, составляющего угол α с положительным направлением оси Ox, к величине ρ этого перемещения, когда ρ стремится к нулю, т. е.*

$$\frac{dz}{d\rho} = \lim_{\rho \to 0} \frac{\Delta z}{\rho} = \lim_{\rho \to 0} \frac{f(x + \Delta x, y + \Delta y) - f(x, y)}{\rho}. \qquad (7.1)$$

Найдем формулу для вычисления производной по направлению α в точке $M(x, y)$ от функции $z = f(x, y)$, дифференцируемой в этой точке. Из рис. 7.1 видно, что $\Delta x = \rho \cos \alpha$, $\Delta y = \rho \sin \alpha$.

Используя эквивалентность полного приращения $\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$ и полного дифференциала $dz = \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y$, из формулы (7.1) получаем

$$\frac{dz}{d\rho} = \lim_{\rho \to 0} \frac{\Delta z}{\rho} = \lim_{\rho \to 0} \frac{dz}{\rho} = \lim_{\rho \to 0} \frac{(\partial z/\partial x)\Delta x + (\partial z/\partial y)\Delta y}{\rho} =$$

$$= \lim_{\rho \to 0} \frac{(\partial z/\partial x)\rho \cos \alpha + (\partial z/\partial y)\rho \sin \alpha}{\rho} = \frac{\partial z}{\partial x} \cos \alpha + \frac{\partial z}{\partial y} \sin \alpha,$$

поскольку значения производных $\partial z/\partial x$, $\partial z/\partial y$ и угол α не зависят от ρ. Итак,

$$\frac{dz}{d\rho} = \frac{\partial z}{\partial x} \cos \alpha + \frac{\partial z}{\partial y} \sin \alpha. \qquad (7.2)$$

Таким образом, производная по направлению α является линейной комбинацией частных производных.

Пример 7.1. Вычислить в точке $M_0(2\sqrt{2}; \sqrt{2})$ производную функции $z = xy + \ln x$ в направлении, составляющем угол $\alpha = \pi/4$ с осью Ox.

Решение. По формуле (7.2) имеем

$$\left.\frac{dz}{d\rho}\right|_{\alpha=\pi/4} = \left(y + \frac{1}{x}\right)_{M_0} \cos \frac{\pi}{4} + (x)_{M_0} \sin \frac{\pi}{4} =$$

$$= \left(\sqrt{2} + \frac{1}{2\sqrt{2}}\right)\frac{\sqrt{2}}{2} + 2\sqrt{2} \cdot \frac{\sqrt{2}}{2} = \frac{13}{4} > 0,$$

т. е. скорость изменения z выражается положительным числом, следовательно, функция z возрастает в точке M_0 в указанном направлении.

Градиент функции и его свойства. В математическом программировании наибольший интерес представляет вопрос о направлении наискорейшего возрастания функции в данной точке. Эта задача решается с помощью вектора, называемого градиентом функции.

Определение 7.2. *Градиентом функции $z = f(x, y)$ в данной точке $M_0(x, y)$ называется вектор, расположенный в плоскости xOy и имеющий своими координатами частные производные функции z, вычисленные в этой точке.*

Градиент обозначают $\operatorname{grad} f(x, y)$, $\mathbf{g}$, $\overline{\nabla f(x, y)}$.

Итак, $\mathbf{g} = (\partial f/\partial x; \partial f/\partial y)$, а модуль градиента

$$|\mathbf{g}| = \mathrm{g} = \sqrt{(\partial f/\partial x)^2 + (\partial f/\partial y)^2}.$$

Теорема 7.1. *Направление градиента является направлением наискорейшего возрастания функции $z = f(x, y)$; модуль градиента равен наибольшей скорости возрастания функции z.*

Доказательство. На рис. 7.1 направление α дифференцирования определяется вектором $\overline{M_0M_1} = \mathbf{a}$. Единичный вектор $\mathbf{a}_0$ этого направления имеет координаты $\cos\alpha$, $\sin\alpha$, ибо

$$|\mathbf{a}_0| = \sqrt{\cos^2\alpha + \sin^2\alpha} = 1.$$

Найдем скалярное произведение векторов $\mathbf{g}$ и $\mathbf{a}_0$:

$$(\mathbf{g}, \mathbf{a}_0) = (\partial f/\partial x)\cos\alpha + (\partial f/\partial y)\sin\alpha.$$

Сравнивая это равенство с равенством (7.2), устанавливаем, что

$$dz/d\rho = (\mathbf{g}, \mathbf{a}_0). \tag{7.3}$$

Обозначим угол между векторами $\mathbf{g}$ и $\mathbf{a}_0$ через φ, тогда

$$(\mathbf{g}, \mathbf{a}_0) = |\mathbf{g}|\,|\mathbf{a}_0|\cos\varphi = \mathrm{g}\cdot 1\cos\varphi.$$

Учитывая это равенство, перепишем соотношение (7.3) в виде

$$dz/d\rho = \mathrm{g}\cos\varphi.$$

Если $\varphi = 0$, т. е. если направление вектора $\mathbf{a}_0$ (направление дифференцирования) совпадает с направлением градиента $\mathbf{g}$, то $\cos\varphi$ достигает своего наибольшего значения (равного 1) и

$$(dz/d\rho)_{\max} = \mathrm{g}. \tag{7.4}$$

Таким образом, производная функции $z = f(x, y)$ достигает наибольшего значения по направлению градиента, что и свидетельствует о том, что в направлении градиента функция z возрастает с наибольшей скоростью. Из формулы (7.4) видно также, что наибольшая скорость возрастания функции z равна модулю градиента. $\triangle$

Пример 7.2. Найти направление наискорейшего возрастания функции $z = x^2 + xy + 5$ в точке $M_0(1; -1)$ и вычислить значение производной в этом направлении.

Решение. Находим координаты градиента данной функции: $\partial z/\partial x = 2x+y$, $\partial z/\partial y = x$. Итак, $\overline{\nabla f(x,y)} = (2x+y; x)$. В точке $M_0(1; -1)$ градиент имеет координаты $\overline{\nabla f(1; -1)} = (1; 1)$. По координатам градиента видно, что искомое направление дифференцирования составляет угол $45°$ с осью Ox. Значение производной в этом направлении

$$(dz/d\rho)_{\max} = \left|\overline{\nabla f(1; -1)}\right| = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

Теорема 7.2. *Градиент функции $z = f(x, y)$ в каждой точке $M_0(x; y)$ направлен по нормали к линии уровня поверхности $z = f(x, y)$, проходящей через эту точку.*

Доказательство. Пусть уравнение линии уровня l (рис. 7.2), проходящей через точку M_0, имеет вид

$$f(x, y) = C. \tag{7.5}$$

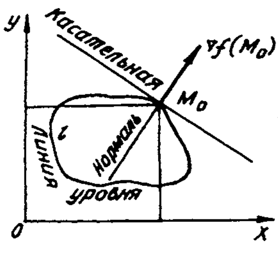


Рис. 7.2

Чтобы найти угловой коэффициент k касательной в точке M_0 к этой линии, продифференцируем равенство (7.5) как неявную функцию:

$$\frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0,$$

откуда

$$dy/dx = k = -\frac{\partial f/\partial x}{\partial f/\partial y}. \tag{7.6}$$

Найдем угловой коэффициент k_1 прямой, на которой лежит градиент $\overline{\nabla f(M_0)} = (\partial f/\partial x; \partial f/\partial y)$. Очевидно, что

$$k_1 = \frac{\partial f/\partial y}{\partial f/\partial x}. \tag{7.7}$$

Сравнивая равенства (7.6) и (7.7), видим, что $kk_1 = -1$, т. е. градиент и касательная взаимно перпендикулярны. Иначе говоря: градиент в данной точке направлен по нормали к линии уровня поверхности. △

Обобщая понятие градиента на случай n-мерного пространства, примем следующее определение.

Определение 7.3. *Если функция* $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ *дифференцируема в точке* $\mathbf{x}_0 = (x_1^0, x_2^0, \ldots, x_n^0)$, *то градиентом* $f(\mathbf{x})$ *в точке* $\mathbf{x}_0$ *называется* n*-мерный вектор, координаты которого равны частным производным функции* $f(\mathbf{x})$, *вычисленным в точке* $\mathbf{x}_0$, т. е.

$$\overline{\nabla f(\mathbf{x}_0)} = (\partial f(\mathbf{x}_0)/\partial x_1; \ldots; \partial f(\mathbf{x}_0)/\partial x_n).$$

Все свойства градиента, установленные в трехмерном пространстве, переносятся и в n-мерное пространство.

Вектор $\overline{-\nabla f(\mathbf{x})}$, противоположный градиенту, называется *антиградиентом.* Он указывает направление наискорейшего убывания функции $f(\mathbf{x})$.

На замечательном свойстве градиента указывать в каждой точке, в которой он существует, направление наискорейшего возрастания функции, основаны градиентные методы решения задач математического программирования, в частности выпуклого программирования.

Выпуклые и вогнутые функции, их основные свойства. Важный класс задач математического программирования составляют задачи, модели которых содержат выпуклые (вогнутые) функции.

Определение 7.4. *Функция* $f(\mathbf{x})$, *определенная на выпуклом множестве* X, *называется выпуклой, если для любых* *точек* $\mathbf{x}'$ *и* $\mathbf{x}''$ *из этого множества и любого* $0 \leqslant \lambda \leqslant 1$ *справедливо неравенство*

$$f(\lambda \mathbf{x}' + (1-\lambda)\mathbf{x}'') \leqslant \lambda f(\mathbf{x}') + (1-\lambda) f(\mathbf{x}''). \tag{7.8}$$

Если в соотношении (7.8) *при* $0 < \lambda < 1$ *и любых* $\mathbf{x}', \mathbf{x}'' \in \in X$ $(\mathbf{x}' \not\equiv \mathbf{x}'')$ *имеет место строгое неравенство, то* $f(\mathbf{x})$ *называется строго выпуклой.*

Определение 7.5. *Функция* $f(\mathbf{x})$, *определенная на выпуклом множестве* X, *называется вогнутой, если для любых точек* $\mathbf{x}'$ *и* $\mathbf{x}''$ *из этого множества и любого* $0 \leqslant \lambda \leqslant 1$ *справедливо неравенство*

$$f(\lambda \mathbf{x}' + (1-\lambda)\mathbf{x}'') \geqslant \lambda f(\mathbf{x}') + (1-\lambda) f(\mathbf{x}''). \tag{7.9}$$

Если в соотношении (7.9) *при* $0 < \lambda < 1$ *и любых* $\mathbf{x}', \mathbf{x}'' \in \in X$ $(\mathbf{x}' \not\equiv \mathbf{x}'')$ *имеет место строгое неравенство, то* $f(\mathbf{x})$ *называется строго вогнутой.*

Геометрическая иллюстрация данных определений для двухмерного пространства ясна из рис. 7.3: выпуклая (рис. 7.3, *а*) функция $f(x)$ на отрезке $[x'; x'']$ не может принимать бóльших значений, чем линейная функция, интерполирующая значения $f(x')$ и $f(x'')$. В свою очередь вогнутая функция $f(x)$ (рис. 7.3, *б*) не может принимать меньших значений, чем линейная функция, интерполирующая значения $f(x')$ и $f(x'')$.

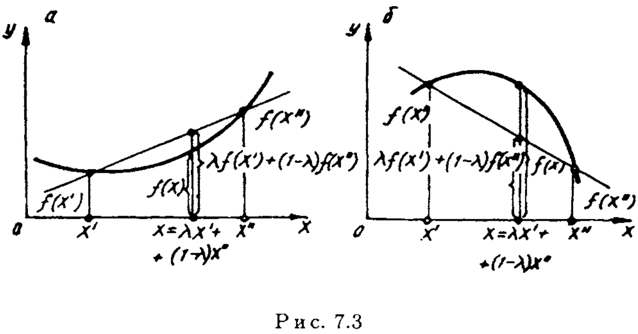


Рис. 7.3

Из определений 7.4 и 7.5 видно, что свойство выпуклости (вогнутости) функций предполагает выпуклость множества, на котором задана функция.

Можно доказать, что если функции $f_i(\mathbf{x})$ $(i = \overline{1,k})$ являются выпуклыми на некотором выпуклом множестве X, то выпуклой на X будет и неотрицательная линейная комбинация этих функций, т. е. функция $f = \sum\limits_{i=1}^{n} \lambda_i f_i(\mathbf{x})$ при $\lambda_i \geqslant 0$, в частности выпуклой будет и сумма выпуклых функций.

Теорема 7.3. *Если $\varphi(\mathbf{x})$ — выпуклая функция при всех $\mathbf{x} \geqslant \mathbf{0}$, то будет выпуклым и множество решений системы $\varphi(\mathbf{x}) \leqslant b,\ \ \mathbf{x} \geqslant \mathbf{0}$.*

Доказательство. Убедимся, что вместе с решениями $\mathbf{x}_1$ и $\mathbf{x}_2$ системы $\varphi(\mathbf{x}) \leqslant b,\ \ \mathbf{x} \geqslant \mathbf{0}$, множеству решений этой системы принадлежит и любой вектор $\mathbf{x}_0 = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, где $0 \leqslant \lambda \leqslant 1$. Ясно, что $\mathbf{x}_0 \geqslant \mathbf{0}$. Остается показать, что $\varphi(\mathbf{x}_0) \leqslant b$. По свойству выпуклости функции $\varphi(\mathbf{x})$

$$\varphi(\mathbf{x}_0) = \varphi(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leqslant \lambda\varphi(\mathbf{x}_1) + (1-\lambda)\varphi(\mathbf{x}_2). \tag{7.10}$$

Но $\varphi(\mathbf{x}_1) \leqslant b,\ \ \varphi(\mathbf{x}_2) \leqslant b$, поэтому для любого $0 \leqslant \lambda \leqslant 1$ справедливы неравенства $\lambda\varphi(\mathbf{x}_1) \leqslant \lambda b$ и $(1-\lambda)\varphi(\mathbf{x}_2) \leqslant (1-\lambda)b$. Складывая эти неравенства, получаем $\lambda\varphi(\mathbf{x}_1) + (1-\lambda)\varphi(\mathbf{x}_2) \leqslant$ $\leqslant \lambda b + (1-\lambda)b = b$. С учетом соотношения (7.10) приходим к неравенству $\varphi(\mathbf{x}_0) \leqslant b$. $\triangle$

Аналогичная теорема доказывается и для вогнутых функций.

Можно показать, что множество решений системы неравенств $\mathbf{x} \leqslant \mathbf{0},\ \ \varphi_i(\mathbf{x})\{\leqslant; \geqslant\}b_i$ $(i = \overline{1,m})$ выпукло, если в неравенствах со знаком $\leqslant$ φ_i — выпуклые функции, а в неравенствах со знаком $\geqslant$ φ_i — вогнутые функции для $\mathbf{x} \geqslant \mathbf{0}$.

Доказывается, что выпуклая функция $f(\mathbf{x})$, определенная на выпуклом множестве X, непрерывна в любой внутренней точке этого множества.

Теорема 7.4. *Если выпуклая функция $f(\mathbf{x})$ дифференцируема во внутренних точках множества X, то для любых внутренних точек $\mathbf{x}_1$ и $\mathbf{x}_2$ имеет место неравенство*

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geqslant \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1). \tag{7.11}$$

Доказательство. Поскольку $f(\mathbf{x})$ — выпуклая функция, то $f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \leqslant (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2)$, откуда

$$\big(f(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)) - f(\mathbf{x}_1)\big)/\lambda \leqslant f(\mathbf{x}_2) - f(\mathbf{x}_1). \tag{7.12}$$

Известно, что если функция $f(x)$ одной переменной x в промежутке от x_0 до $x_0 + \Delta x$ непрерывна и имеет непрерывную производную, то справедлива следующая формула Тейлора:

$$f(x_0 + \Delta x) = f(x_0) + \Delta x \, \frac{df(x_0 + \theta \Delta x)}{dx} \quad (0 < \theta < 1).$$

В нашем случае речь идет о функции $f(\mathbf{x})$, зависящей от n переменных, поэтому формула Тейлора в векторной записи принимает следующий вид:

$$f(\mathbf{x}_0 + \Delta\mathbf{x}) = f(\mathbf{x}_0) + \Delta\mathbf{x} \, \nabla f(\mathbf{x}_0 + \theta\Delta\mathbf{x}) \quad (0 < \theta < 1).$$

Используя ее при $\mathbf{x}_0 = \mathbf{x}_1$, $\Delta\mathbf{x} = \lambda(\mathbf{x}_2 - \mathbf{x}_1)$, будем иметь

$$f(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)) = f(\mathbf{x}_1) +$$
$$+ \lambda(\mathbf{x}_2 - \mathbf{x}_1)\nabla f(\mathbf{x}_1 + \theta\lambda(\mathbf{x}_2 - \mathbf{x}_1)). \tag{7.13}$$

Подставляя выражение (7.13) в неравенство (7.12), находим

$$(\mathbf{x}_2 - \mathbf{x}_1)\nabla f(\mathbf{x}_1 + \theta\lambda(\mathbf{x}_2 - \mathbf{x}_1)) \leqslant f(\mathbf{x}_2) - f(\mathbf{x}_1).$$

Переходя к пределу при $\lambda \to 0$ в этом неравенстве, получаем соотношение (7.11). $\triangle$

Теорема 7.5. *Выпуклая функция $f(\mathbf{x})$, определенная на выпуклом множестве X, достигает своего глобального минимума в каждой точке $\mathbf{x}$, в которой градиент функции обращается в нуль.*

Доказательство. Пусть, например, в точке $\mathbf{x}_0$ градиент $\nabla f(\mathbf{x}_0) = \mathbf{0}$. Покажем, что в этой точке функция $f(\mathbf{x})$ имеет глобальный минимум, т. е. что для любой точки $\mathbf{x} \in X$ справедливо неравенство $f(\mathbf{x}) \geqslant f(\mathbf{x}_0)$. Известно, что для выпуклой на множестве X функции $f(\mathbf{x})$ справедливо неравенство (7.11), где $\mathbf{x}_1$ и $\mathbf{x}_2$ — любые внутренние точки множества X. Полагая в неравенстве (7.11) $\mathbf{x}_1 = \mathbf{x}_0$ и $\mathbf{x}_2 = \mathbf{x}$, получаем $f(\mathbf{x}) - f(\mathbf{x}_0) \geqslant \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$, откуда $f(\mathbf{x}) \geqslant f(\mathbf{x}_0)$, что и доказывает теорему. $\triangle$

Теорема 7.6. *Локальный минимум выпуклой функции $f(\mathbf{x})$, определенной на выпуклом множестве X, совпадает с ее глобальным минимумом на этом множестве.*

Доказательство. Пусть функция $f(\mathbf{x})$ в точке $\mathbf{x}_0$ имеет локальный минимум, т. е. для всех точек $\mathbf{x}$ из некоторой ε-окрестности точки $\mathbf{x}_0$ справедливо неравенство

$$f(\mathbf{x}_0) \leqslant f(\mathbf{x}). \tag{7.14}$$

Докажем, что неравенство (7.14) справедливо для любой точки допустимой области X. Предположим противное, т. е. что в области X существует точка $\mathbf{x}'$, такая, что

$$f(\mathbf{x}') < f(\mathbf{x}_0), \tag{7.15}$$

и, следовательно, $\mathbf{x}_0$ не является точкой глобального минимума.

Рассмотрим отрезок, соединяющий точки $\mathbf{x}'$ и $\mathbf{x}_0$. Его уравнение имеет вид $\mathbf{x} = \lambda\mathbf{x}' + (1-\lambda)\mathbf{x}_0$ $(0 \leqslant \lambda \leqslant 1)$. Ввиду того что допустимая область X выпуклая, отрезок $\mathbf{x}'\mathbf{x}_0$ полностью лежит в ней. Возьмем на нем точку $\mathbf{x}''$, лежащую в ε-окрестности точки $\mathbf{x}_0$, но не совпадающую с точкой $\mathbf{x}_0$. Поскольку точка $\mathbf{x}''$ расположена на отрезке $\mathbf{x}'\mathbf{x}_0$, ей отвечает некоторая определенная величина $\lambda = \lambda''$ и выполняется условие $\mathbf{x}'' = \lambda''\mathbf{x}' + (1-\lambda'')\mathbf{x}_0$ $(0 < \lambda'' < 1)$. В силу выпуклости функции $f(\mathbf{x})$ для точек $\mathbf{x}'$ и $\mathbf{x}_0$ имеет место неравенство $f(\lambda''\mathbf{x}' + (1-\lambda'')\mathbf{x}_0) \leqslant \lambda'' f(\mathbf{x}') + (1-\lambda'')f(\mathbf{x}_0)$ или

$$f(\mathbf{x}'') \leqslant f(\mathbf{x}_0) + \lambda''(f(\mathbf{x}') - f(\mathbf{x}_0)). \tag{7.16}$$

Но по предположению выполняется неравенство (7.15), а поэтому в соотношении (7.16) $\lambda''(f(\mathbf{x}') - f(\mathbf{x}_0)) < 0$. Отбрасывая это слагаемое, мы можем только усилить неравенство (7.16), т. е. получаем

$$f(\mathbf{x}'') < f(\mathbf{x}_0). \tag{7.17}$$

Однако неравенство (7.17) противоречит условию (7.14) для точек ε-окрестности точки $\mathbf{x}_0$. Значит, в допустимой области X нет точек типа $\mathbf{x}'$, для которых выполняется условие (7.15). Следовательно, точка $\mathbf{x}_0$ действительно является точкой глобального минимума. $\triangle$

Для практической проверки функции на выпуклость может оказаться полезным следующее свойство выпуклых функций: функция $f(\mathbf{x})$ будет выпуклой, если ее вторые частные производные образуют матрицу, в которой все главные миноры неотрицательны.

7.2. ЗАДАЧА ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

Задача математического программирования

$$\max(\min) z = f(\mathbf{x}); \tag{7.18}$$

$$\left.\begin{array}{r} \varphi_i(\mathbf{x}) \{\leqslant, =, \geqslant\}\, b_i \quad (i = \overline{1, m}\,), \\ \mathbf{x} \geqslant 0, \quad \mathbf{x} = (x_1, \ldots, x_n), \end{array}\right\} \tag{7.19}$$

в которой либо целевая функция (7.18), либо ограничения (7.19), либо и то и другое нелинейны, называется *нелинейной*.

Нелинейные задачи составляют широкий класс настолько сложных задач, что до сих пор невозможно разработать общие методы, подобные симплекс-методу в линейном программировании, которые позволяли бы решать любые нелинейные задачи. Но, несмотря на отсутствие универсальных методов, разработаны способы решения отдельных специальных классов задач, и прежде всего задач с выпуклыми (вогнутыми) функциями $f(\mathbf{x})$ и $\varphi_i(\mathbf{x})$.

В теории выпуклого программирования в качестве основной обычно рассматривается задача минимизации выпуклой функции n переменных $z = f(\mathbf{x})$ при ограничениях $\varphi_i(\mathbf{x}) \leqslant 0$ $(i = \overline{1, m}\,)$, $\mathbf{x} \geqslant \mathbf{0}$, где функции $\varphi_i(\mathbf{x})$ предполагаются выпуклыми. Выпуклость множества допустимых решений задачи следует из теоремы 7.3.

Если $f(\mathbf{x})$ и $\varphi_i(\mathbf{x})$ являются вогнутыми функциями, то имеем задачу максимизации $f(\mathbf{x})$ при ограничениях $\varphi_i(\mathbf{x}) \geqslant 0$ $(i = \overline{1, m}\,)$, $\mathbf{x} \geqslant \mathbf{0}$.

7.3. МЕТОД МНОЖИТЕЛЕЙ ЛАГРАНЖА. ЭКОНОМИЧЕСКИЙ СМЫСЛ МНОЖИТЕЛЕЙ ЛАГРАНЖА

Метод множителей Лагранжа является классическим методом решения задач математического программирования (в частности выпуклого). К сожалению, при практическом применении метода могут встретиться значительные вычислительные трудности, сужающие область его использования.

Мы рассматриваем здесь метод Лагранжа главным образом потому, что он является аппаратом, активно используемым для обоснования различных современных численных методов, широко применяемых на практике. Что же касается функции Лагранжа и множителей Лагранжа, то они играют самостоятельную и исключительно важную роль в теории и приложениях не только математического программирования.

Рассмотрим классическую задачу оптимизации

$$\max(\min) z = f(\mathbf{x}); \tag{7.20}$$

$$\varphi_i(\mathbf{x}) = b_i \quad (i = \overline{1,m}), \qquad \mathbf{x} = (x_1, \dots, x_n). \tag{7.21}$$

Эта задача выделяется из задачи (7.18), (7.19) тем, что среди ограничений (7.21) нет неравенств, нет условий неотрицательности переменных, их дискретности, $m < n$ и функции $f(\mathbf{x})$ и $\varphi_i(\mathbf{x})$ непрерывны и имеют частные производные по крайней мере второго порядка.

Классический подход к решению задачи (7.20), (7.21) дает систему уравнений (необходимые условия), которым должна удовлетворять точка $\mathbf{x}^*$, доставляющая функции $f(\mathbf{x})$ локальный экстремум на множестве точек, удовлетворяющих ограничениям (7.21) (для задачи выпуклого программирования найденная точка $\mathbf{x}^*$ в соответствии с теоремой 7.6 будет одновременно и точкой глобального экстремума).

Предположим, что в точке $\mathbf{x}^*$ функция (7.20) имеет локальный условный экстремум и ранг матрицы $[\partial\varphi_i/\partial x_j]_{m\times n}$ равен m. Тогда необходимые условия запишутся в виде:

$$\left.\begin{aligned} \frac{\partial L}{\partial x_j} &= \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i}{\partial x_j} = 0 \quad (j = \overline{1,n}), \\ \frac{\partial L}{\partial \lambda_i} &= b_i - \varphi_i = 0 \quad (i = \overline{1,m}), \end{aligned}\right\} \tag{7.22}$$

где

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) + \\ + \sum_{i=1}^{m} \lambda_i (b_i - \varphi_i(x_1, \dots, x_n)) \tag{7.23}$$

есть функция Лагранжа; $\lambda_1, \dots, \lambda_m$ — множители Лагранжа.

Существуют также и достаточные условия, при выполнении которых решение системы уравнений (7.22) определяет точку экстремума функции $f(\mathbf{x})$. Этот вопрос решается на основании исследования знака второго дифференциала функции Лагранжа. Однако достаточные условия представляют главным образом теоретический интерес.

Можно указать следующий порядок решения задачи (7.20), (7.21) методом множителей Лагранжа:

1) составить функцию Лагранжа (7.23);

2) найти частные производные функции Лагранжа по всем переменным $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$ и приравнять их нулю. Тем самым будет получена система (7.22), состоящая из $n+m$ уравнений. Решить полученную систему (если это окажется возможным!) и найти таким образом все стационарные точки функции Лагранжа;

3) из стационарных точек, взятых без координат $\lambda_1, \ldots, \lambda_m$, выбрать точки, в которых функция $f(\mathbf{x})$ имеет условные локальные экстремумы при наличии ограничений (7.21). Этот выбор осуществляется, например, с применением достаточных условий локального экстремума. Часто исследование упрощается, если использовать конкретные условия задачи.

Пример 7.3. Найти оптимальное распределение ограниченного ресурса в a единиц между n потребителями, если прибыль, получаемая при выделении j-му потребителю x_j единиц ресурса, вычисляется по формуле $c_j\sqrt{x_j}$.

Решение. Математическая модель задачи имеет следующий вид:

$$\max z = \sum_{j=1}^{n} c_j\sqrt{x_j};$$

$$\sum_{j=1}^{n} x_j = a.$$

Составляем функцию Лагранжа:

$$L(x_1, \ldots, x_n, \lambda) = \sum_{j=1}^{n} c_j\sqrt{x_j} + \lambda\Big(a - \sum_{j=1}^{n} x_j\Big).$$

Находим частные производные функции Лагранжа и приравниваем их нулю:

$$\left.\begin{aligned} &\partial L/\partial x_j = c_j/2\sqrt{x_j} - \lambda = 0 \quad (j = \overline{1,n}),\\ &\partial L/\partial \lambda = a - \sum_{j=1}^{n} x_j = 0. \end{aligned}\right\}$$

Решая эту систему уравнений, получаем:

$$x_j^* = ac_j^2 \Big/ \sum_{j=1}^{n} c_j^2 \quad (j = \overline{1,n}), \qquad \lambda^* = \sqrt{\sum_{j=1}^{n} c_j^2 \Big/ (4a)}.$$

Таким образом, если j-му ($j = \overline{1,n}$) потребителю будет выделено $ac_j^2 \Big/ \sum\limits_{j=1}^{n} c_j^2$ единиц ресурса, то суммарная прибыль достигнет максимальной величины и составит $\sum\limits_{j=1}^{n} c_j \sqrt{x_j^*}$ ден. ед.

Мы рассмотрели метод Лагранжа применительно к классической задаче оптимизации. Можно обобщить этот метод на случай, когда переменные неотрицательны и некоторые ограничения заданы в форме неравенств. Однако это обобщение имеет преимущественно теоретическое значение и не приводит к конкретным вычислительным алгоритмам.

В заключение дадим множителям Лагранжа экономическую интерпретацию. Для этого обратимся к простейшей классической задаче оптимизации

$$\max(\min) z = f(x_1, x_2); \tag{7.24}$$

$$\varphi(x_1, x_2) = b. \tag{7.25}$$

Предположим, что условный экстремум достигается в точке $\mathbf{x}^* = (x_1^*, x_2^*)$. Соответствующее экстремальное значение функции $f(\mathbf{x})$

$$f^* = f(x_1^*, x_2^*).$$

Допустим, что в ограничениях (7.25) величина b может меняться. Тогда координаты x_1^* и x_2^* точки экстремума, а следовательно, и экстремальное значение f^* функции $f(\mathbf{x})$ станут величинами, зависящими от b, т.е. $x_1^* = x_1^*(b)$, $x_2^* = x_2^*(b)$, $f^* = f(x_1^*(b), x_2^*(b)))$. Поэтому производная функции (7.24)

$$df^*/db = \partial f/\partial x_1 \cdot dx_1^*/db + \partial f/\partial x_2 \cdot dx_2^*/db. \tag{7.26}$$

С другой стороны, в силу равенства (7.25) $\varphi(x_1^*(b), x_2^*(b)) = b$, откуда после дифференцирования имеем

$$\partial\varphi/\partial x_1 \cdot dx_1^*/db + \partial\varphi/\partial x_2 \cdot dx_2^*/db = 1. \tag{7.27}$$

Кроме того, в точке экстремума $\mathbf{x}^*$ выполняются необходимые условия (7.22). Из этих равенств для $n = 2$ и $m = 1$ получаем:

$$\partial f/\partial x_1 = \lambda \partial\varphi/\partial x_1, \quad \partial f/\partial x_2 = \lambda \partial\varphi/\partial x_2. \tag{7.28}$$

Подставляя выражения (7.28) в равенство (7.26) и учитывая соотношение (7.27), находим $df^*/db = \lambda\partial\varphi/\partial x_1 \cdot dx_1^*/db + \lambda\partial\varphi/\partial x_2 \cdot dx_2^*/db = \lambda \cdot 1$ или $df^*/db = \lambda$. Для задачи (7.20), (7.21) аналогично получаем $df^*/db = \lambda_i$ $(i = \overline{1,m})$.

Если f интерпретировать как доход или стоимость, а b_i — как объемы некоторых ресурсов, то множители Лагранжа λ_i показывают, как изменится максимальный доход (или минимальная стоимость), если количество ресурса i-го вида увеличится на единицу.

7.4. ГРАДИЕНТНЫЕ МЕТОДЫ

Задачи без ограничений. Градиентным методом можно решать, вообще говоря, любую нелинейную задачу. Однако при этом находится лишь локальный экстремум. Поэтому целесообразнее применять этот метод при решении задач выпуклого программирования, в которых любой локальный экстремум является одновременно и глобальным (см. теорему 7.6).

Будем рассматривать задачу максимизации нелинейной дифференцируемой функции $f(\mathbf{x})$. Суть градиентного поиска точки максимума $\mathbf{x}^*$ весьма проста: надо взять произвольную точку $\mathbf{x}_0$ и с помощью градиента $\nabla f(\mathbf{x}_0)$, вычисленного в этой точке, определить направление, в котором $f(\mathbf{x})$ возрастает с наибольшей скоростью (рис. 7.4), а затем, сделав небольшой шаг в найденном направлении, перейти в новую точку $\mathbf{x}_1$. Потом снова определить наилучшее направление $\nabla f(\mathbf{x}_1)$ для перехода в очередную точку $\mathbf{x}_2$ и т. д. На рис. 7.4 поисковая траектория представляет собой ломаную $\mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\ldots$ Таким образом, надо построить последовательность точек $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots$ так, чтобы она сходилась к точке максимума $\mathbf{x}^*$, т. е. для точек последовательности выполнялись условия

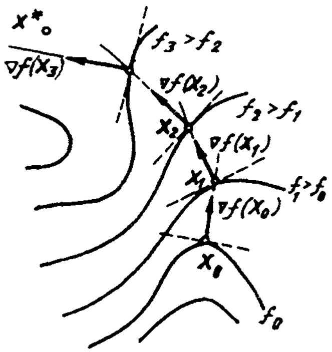


Рис. 7.4

$$f_0 < f_1 < f_2 < \ldots < f_k < \ldots$$

Градиентные методы, как правило, позволяют получать точное решение за бесконечное число шагов и только в некоторых случаях — за конечное. В связи с этим градиентные методы относят к приближенным методам решения.

Движение из точки $\mathbf{x}_k$ в новую точку $\mathbf{x}_{k+1}$ осуществляется по прямой, проходящей через точку $\mathbf{x}_k$ и имеющей уравнение

$$\mathbf{x} = \mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k), \tag{7.29}$$

где λ_k — числовой параметр, от которого зависит величина шага. Как только значение параметра в уравнении (7.29) выбрано: $\lambda_k = \lambda_k^0$, становится определенной очередная точка $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k^0 \nabla f(\mathbf{x}_k)$ на поисковой ломаной.

Градиентные методы отличаются друг от друга способом выбора величины шага — значения λ_k^0 параметра λ_k. Можно, например, двигаться из точки в точку с постоянным шагом $\lambda_k = \lambda$, т. е. при любом k

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda \nabla f(\mathbf{x}_k).$$

Если при этом окажется, что $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$, то следует возвратиться в точку $\mathbf{x}_k$ и уменьшить значение параметра, например до $\lambda/2$.

Иногда величина шага берется пропорциональной модулю градиента.

Если ищется приближенное решение, то поиск можно прекратить, основываясь на следующих соображениях. После каждой серии из определенного числа шагов сравнивают достигнутые значения целевой функции $f(\mathbf{x})$. Если после очередной серии изменение $f(\mathbf{x})$ не превышает некоторого наперед заданного малого числа δ, поиск прекращают и достигнутое значение $f(\mathbf{x})$ рассматривают как искомый приближенный максимум, а соответствующее ему $\mathbf{x}$ принимают за $\mathbf{x}^*$.

Если целевая функция $f(\mathbf{x})$ вогнутая (выпуклая), то необходимым и достаточным условием оптимальности точки $\mathbf{x}^*$ является равенство нулю градиента функции в этой точке.

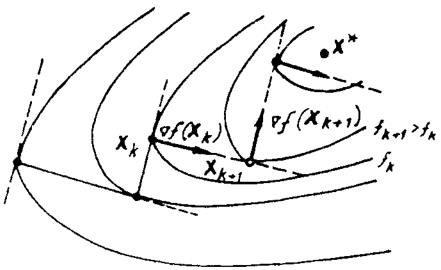
Распространенным является вариант градиентного поиска, называемый *методом наискорейшего подъема*. Суть его в следующем. После определения градиента $\nabla f(\mathbf{x}_k)$ в точке $\mathbf{x}_k$ движение вдоль прямой $\mathbf{x} = \mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k)$ производится до точки $\mathbf{x}_{k+1}$, в которой достигается максимальное значение функции $f(\mathbf{x})$ в направлении градиента $\nabla f(\mathbf{x}_k)$. Затем в этой точке вновь определяется градиент, и движение совершается по прямой $\mathbf{x} = \mathbf{x}_{k+1} + \lambda_{k+1} \nabla f(\mathbf{x}_{k+1}))$ в направлении нового градиента $\nabla f(\mathbf{x}_{k+1})$ до точки $\mathbf{x}_{k+2}$, в которой достигается максимальное в этом направлении значение $f(\mathbf{x})$. Движение продолжается до тех пор, пока не будет достигнута точка $\mathbf{x}^*$, соответствующая наибольшему значению целевой функции $f(\mathbf{x})$. На рис. 7.5 приведена схема движения к оптимальной точке $\mathbf{x}^*$ методом наискорейшего подъема. В данном случае направление градиента $\nabla f(\mathbf{x}_k)$ в точке $\mathbf{x}_k$ является касательным к линии уровня поверхности $f(\mathbf{x})$ в точке $\mathbf{x}_{k+1}$, следовательно, градиент $\nabla f(\mathbf{x}_{k+1})$ в точке $\mathbf{x}_{k+1}$ ортогонален градиенту $\nabla f(\mathbf{x}_k)$ (сравните с рис. 7.4).

Перемещение из точки $\mathbf{x}_k$ в точку $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k)$ сопровождается возрастанием функции $f(\mathbf{x})$ на величину

$$\Delta f(\mathbf{x}) = f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) = f(x_{k+1,1}; \ldots ; x_{k+1,n}) -$$

$$-f(x_{k1}; \ldots ; x_{kn}) = f(x_{k1} + \lambda_k \partial f(\mathbf{x}_k)/\partial x_1; \ldots ;$$

$$x_{kn} + \lambda_k \partial f(\mathbf{x}_k)/\partial x_n) - f(x_{k1}; \ldots ; x_{kn}). \qquad (7.30)$$



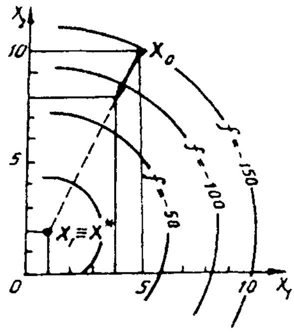


Рис. 7.5 Рис. 7.6

Из выражения (7.30) видно, что приращение Δf является функцией переменной λ_k, т. е. $\Delta f = \Delta f(\lambda_k)$. При нахождении максимума функции $f(\mathbf{x})$ в направлении градиента $\nabla f(\mathbf{x}_k)$ необходимо выбирать шаг перемещения (множитель λ_k), обеспечивающий наибольшее возрастание приращению функции, именно функции $\Delta f(\lambda_k)$. Величина λ_k, при которой достигается наибольшее значение $\Delta f(\lambda_k)$, может быть определена из необходимого условия экстремума функции $\Delta f(\lambda_k)$:

$$d(\Delta f(\lambda_k))/d\lambda_k = 0. \tag{7.31}$$

Найдем выражение для производной, дифференцируя равенство (7.30) по λ_k как сложную функцию:

$$d(\Delta f(\lambda_k))/d\lambda_k = \partial f(x_{k+1})/\partial x_1 \cdot \partial f(x_k)/\partial x_1 + \ldots +$$
$$+\partial f(x_{k+1})/\partial x_n \cdot \partial f(x_k)/\partial x_n = \nabla f(\mathbf{x}_{k+1}) \cdot \nabla f(\mathbf{x}_k).$$

Подставляя этот результат в равенство (7.31), получаем

$$d(\Delta f(\lambda_k))/d\lambda_k = \nabla f(\mathbf{x}_{k+1}) \cdot \nabla f(\mathbf{x}_k) = 0. \tag{7.32}$$

Это равенство имеет простое геометрическое истолкование: градиент в очередной точке $\mathbf{x}_{k+1}$ ортогонален градиенту в предыдущей точке $\mathbf{x}_k$.

Пример 7.4. Определить максимум функции $f = 4x_1 + 8x_2 - 2x_1^2 - 2x_2^2$, начав оптимизационный поиск с точки $\mathbf{x}_0 = (5; 10)$.

Решение. Будем сопровождать решение задачи графической иллюстрацией (рис. 7.6). По данному уравнению параболоида вращения построены линии уровня этой поверхности, для чего уравнение приведено к виду $(x_1-1)^2+(x_2-2)^2=5-0,5f$, из которого ясно, что линиями пересечения параболоида с плоскостями, параллельными плоскости x_1Ox_2 (линиями уровня), являются окружности радиусом $\sqrt{5-0,5f}$. При $f = -150, -100, -50$ их радиусы равны соответственно $\sqrt{80}$, $\sqrt{55}$, $\sqrt{30}$, а общий центр находится в точке (1; 2).

Находим градиент данной функции:

$$\nabla f(\mathbf{x}) = \nabla f(x_1, x_2) = (\partial f/\partial x_1; \partial f/\partial x_2) = (-4x_1+4; -4x_2+8).$$

I шаг. Вычисляем:

$$\nabla f(\mathbf{x}_0) = \nabla f(5;10) = (-4\cdot 5+4; -4\cdot 10+8) = (-16; -32).$$

На рис. 7.6 в точке $\mathbf{x}_0 = (5;10)$ построен вектор $1/16\nabla f(\mathbf{x}_0) = (-1;-2)$, указывающий направление наискорейшего возрастания функции в точке $\mathbf{x}_0$. На этом направлении расположена следующая точка $\mathbf{x}_1 = \mathbf{x}_0+\lambda_0\nabla f(\mathbf{x}_0) = (5;10)+\lambda_0(-16;-32) = (5-16\lambda_0; 10-32\lambda_0)$. В этой точке

$$\nabla f(\mathbf{x}_1) = (-4(5-16\lambda_0)+4; -4(10-32\lambda_0)+8 =$$
$$= (-16+64\lambda_0; -32+128\lambda_0).$$

Используя условие (7.32), получаем

$$d(\Delta f(\lambda_0))/d\lambda_0 = \nabla f(\mathbf{x}_1)\cdot\nabla f(\mathbf{x}_0) =$$
$$= (-16+64\lambda_0; -32+128\lambda_0)(-16;-32) = 0$$

или $1-4\lambda_0 = 0$, откуда $\lambda_0 = 1/4$. Так как $d^2(\Delta f(\lambda_0))/d\lambda_0^2 < 0$, то найденное значение λ_0 является точкой максимума $\Delta f(\mathbf{x})$. Находим $\mathbf{x}_1 = (5-16/4; 10-32/4) = (1;2)$.

II шаг. Начальная точка для второго шага $\mathbf{x}_1 = (1;2)$. Вычисляем $\nabla f(\mathbf{x}_1) = (-4\cdot 1+4; -4\cdot 2+8) = (0;0)$. Следовательно, $\mathbf{x}_1 = (1;2)$ является стационарной точкой. Но поскольку данная функция вогнутая, то в найденной точке (1; 2) достигается глобальный максимум $f_{\max} = 10$.

Задача с линейными ограничениями. Сразу же отметим, что если целевая функция $f(\mathbf{x})$ в задаче с ограничениями имеет единственный экстремум и он находится внутри допустимой области, то для поиска экстремальной точки $\mathbf{x}^*$ применяется изложенная выше методика без каких-либо изменений.

Рассмотрим задачу выпуклого программирования с линейными ограничениями:

$$\max z = f(\mathbf{x}); \tag{7.33}$$

$$\mathbf{a}_i\mathbf{x} \leqslant a_{i0} \quad (i=\overline{1,m}), \tag{7.34}$$

$$\mathbf{x} \geqslant \mathbf{0}, \tag{7.35}$$

где $\mathbf{x} = (x_1; \ldots; x_n)$; $\mathbf{a}_i = (a_{i1}; \ldots; a_{in})$.

Предполагается, что $f(\mathbf{x})$ является вогнутой функцией и имеет непрерывные частные производные в каждой точке допустимой области.

Начнем с геометрической иллюстрации процесса решения задачи (рис. 7.7). Пусть начальная точка $\mathbf{x}_0$ расположена внутри допустимой области. Из точки $\mathbf{x}_0$ можно двигаться в направлении градиента $\nabla f(\mathbf{x}_0)$, пока $f(\mathbf{x})$ не достигнет максимума. В нашем случае $f(\mathbf{x})$ все время возрастает, поэтому остановиться надо в точке $\mathbf{x}_1$ на граничной прямой. Как

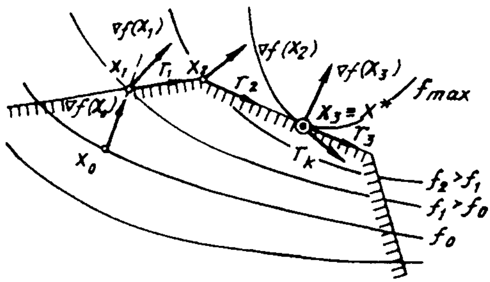


Рис. 7.7

видно из рисунка, дальше двигаться в направлении градиента $\nabla f(\mathbf{x}_1)$ нельзя, так как выйдем из допустимой области. Поэтому надо найти другое направление перемещения, которое, с одной стороны, не выводит из допустимой области, а с другой — обеспечивает наибольшее возрастание $f(\mathbf{x})$. Такое направление определит вектор $\mathbf{r}_1$, составляющий с вектором $\nabla f(\mathbf{x}_1)$ наименьший острый угол по сравнению с любым другим вектором, выходящим из точки $\mathbf{x}_1$ и лежащим в допустимой области. Аналитически такой вектор найдется из условия максимизации скалярного произведения $\nabla f(\mathbf{x}_1) \cdot \mathbf{r}_1 > 0$. В данном случае вектор $\mathbf{r}_1$, указывающий наивыгоднейшее направление, совпадает с граничной прямой.

Таким образом, на следующем шаге двигаться надо по граничной прямой до тех пор, пока возрастает $f(\mathbf{x})$; в нашем случае — до точки $\mathbf{x}_2$. Из рисунка видно, что далее следует перемещаться в направлении вектора $\mathbf{r}_2$, который находится из условия максимизации скалярного произведения $\nabla f(\mathbf{x}_2) \cdot \mathbf{r}_2 > 0$, т. е. по граничной прямой. Движение заканчивается в точке $\mathbf{x}_3$, поскольку в этой точке завершается оптимизационный поиск, ибо в ней функция $f(\mathbf{x})$ имеет локальный максимум. Ввиду вогнутости в этой точке $f(\mathbf{x})$ достигает также глобального максимума в допустимой области. Градиент в точке максимума $\mathbf{x}_3 = \mathbf{x}^*$ составляет тупой угол с любым вектором $\mathbf{r}_k$ из допустимой области, проходящим через $\mathbf{x}_3$, поэтому скалярное произведение $\nabla f(\mathbf{x}_3) \cdot \mathbf{r}_k$ будет отрицательным для любого допустимого $\mathbf{r}_k$, кроме $\mathbf{r}_3$, направленного по граничной прямой. Для него скалярное произведение $\nabla f(\mathbf{x}_3) \cdot \mathbf{r}_3 = 0$, так как $\nabla f(\mathbf{x}_3)$ и $\mathbf{r}_3$ взаимно перпендикулярны (граничная прямая касается линии уровня поверхности $f(\mathbf{x})$, проходящей через точку максимума $\mathbf{x}^*$). Это равенство и служит аналитическим признаком того, что в точке $\mathbf{x}_3$ функция $f(\mathbf{x})$ достигла максимума.

Рассмотрим теперь аналитическое решение задачи (7.33) — (7.35). Если оптимизационный поиск начинается с точки, лежащей в допустимой области (все ограничения задачи выполняются как строгие неравенства), то перемещаться следует по направлению градиента так, как установлено выше. Однако теперь выбор λ_k в уравнении (7.29) усложняется требованием, чтобы очередная точка $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k)$ оставалась в допустимой области. Это означает, что ее координаты должны удовлетворять ограничениям (7.34), (7.35), т. е. должны выполняться неравенства:

$$\left.\begin{aligned} \mathbf{a}_i(\mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k)) \leqslant a_{i0} \quad (i = \overline{1, m}), \\ \mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k) \geqslant 0. \end{aligned}\right\} \tag{7.36}$$

Решая систему линейных неравенств (7.36), находим отрезок $[\lambda_k'; \lambda_k'']$ допустимых значений параметра λ_k, при которых точка $\mathbf{x}_{k+1}$ будет принадлежать допустимой области.

Значение λ_k^*, определяемое в результате решения уравнения (7.32): $\nabla f(\mathbf{x}_k + \lambda_k \nabla f(\mathbf{x}_k)) \cdot \nabla f(\mathbf{x}_k) = 0$, при котором $f(\mathbf{x})$ имеет локальный максимум по λ_k в направлении $\nabla f(\mathbf{x}_k)$, должно принадлежать отрезку $[\lambda_k'; \lambda_k'']$. Если же найденное значение λ_k выходит за пределы указанного отрезка, то в качестве λ_k^* принимается λ_k''. В этом случае очередная точка поисковой траектории оказывается на граничной гиперплоскости, соответствующей тому неравенству системы (7.36), по которому при решении системы получена правая конечная точка λ_k'' отрезка допустимых значений параметра λ_k.

Если оптимизационный поиск начат с точки, лежащей на граничной гиперплоскости, или очередная точка поисковой траектории оказалась на граничной гиперплоскости, то для продолжения движения к точке максимума прежде всего необходимо найти наилучшее направление движения. С этой целью следует решить вспомогательную задачу математического программирования, а именно: максимизировать функцию

$$T_k = \nabla f(\mathbf{x}_k) \cdot \mathbf{r}_k \tag{7.37}$$

при ограничениях

$$\mathbf{a}_i \cdot \mathbf{r}_k \leqslant 0 \tag{7.38}$$

для тех i, при которых

$$\mathbf{a}_i \cdot \mathbf{x}_k = a_{i0}, \tag{7.39}$$

$$|\mathbf{r}_k| = 1, \tag{7.40}$$

где $\mathbf{r}_k = (r_{k1}; \ldots ; r_{kn})$; $|\mathbf{r}_k| = \sqrt{\sum\limits_{j=1}^{n} r_{kj}^2}$.

В результате решения задачи (7.37) — (7.40) будет найден вектор $\mathbf{r}_k$, составляющий с градиентом наименьший острый угол.

Условие (7.39) говорит о том, что точка $\mathbf{x}_k$ принадлежит границе допустимой области, а условие (7.38) означает, что перемещение из $\mathbf{x}_k$ по вектору $\mathbf{r}_k$ будет направлено внутрь допустимой области или по ее границе. Условие нормализации (7.40) необходимо для ограничения величины $\mathbf{r}_k$, так как в противном случае значение целевой функции (7.37) можно сделать сколь угодно большим. Известны различные формы условий нормализации, и в зависимости от этого задача (7.37) — (7.40) может быть линейной или нелинейной.

После определения направления $\mathbf{r}_k$ находится значение λ_k^* для следующей точки $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k^* \mathbf{r}_k$ поисковой траектории. При этом используется необходимое условие экстремума в форме, аналогичной уравнению (7.32), но с заменой $\nabla f(\mathbf{x}_k)$ на вектор $\mathbf{r}_k$, т. е.

$$\nabla f(\mathbf{x}_{k+1}) \cdot \mathbf{r}_k = 0. \tag{7.41}$$

Оптимизационный поиск прекращается, когда достигнута точка $\mathbf{x}_k^*$, в которой $\max T_k = \nabla f(\mathbf{x}_k^*) \cdot \mathbf{r}_k = 0$.

Пример 7.5. Максимизировать функцию $f = x_1 + 2x_2 - 0,2x_1^2 - 0,2x_2^2$ при ограничениях $x_1 + 4x_2 \leqslant 14$, $7x_1 + 3x_2 \leqslant 42$, $x_1 \geqslant 0$, $x_2 \geqslant 0$.

Решение. Для наглядного представления процесса оптимизации будем сопровождать его графической иллюстрацией. На рис. 7.8 изображено несколько линий уровня данной поверхности и допустимая область $OABC$, в которой следует найти точку $\mathbf{x}^*$, доставляющую максимум данной функции (см. пример 7.4).

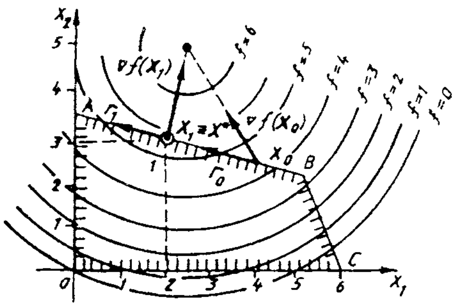


Рис. 7.8

Начнем оптимизационный поиск, например, с точки $\mathbf{x}_0 = (4; 2,5)$, лежащей на граничной прямой AB: $x_1 + 4x_2 = 14$. При этом $f(\mathbf{x}_0) = 4,55$.

Найдем значение градиента $\nabla f(\mathbf{x}) = (1 - 0,4x_1; 2 - 0,4x_2)$ в точке $\mathbf{x}_0$: $\nabla f(\mathbf{x}_0) = (-0,6; 1) \neq (0; 0)$. Кроме того, и из рисунка видно, что через допустимую область проходят линии уровня с пометками более высокими, чем $f(\mathbf{x}_0) = 4,55$. Словом, надо искать направление $\mathbf{r}_0 = (r_{01}; r_{02})$ перемещения в следующую точку $\mathbf{x}_1$, более близкую к оптимальной. С этой целью решаем задачу (7.37) — (7.40) максимизации функции $T_0 = \nabla f(\mathbf{x}_0) \cdot \mathbf{r}_0 = (-0,6; 1)(r_{01}; r_{02}) = -0,6r_{01} + r_{02}$ при следующих ограничениях:

$$\mathbf{a}_1 \cdot \mathbf{r}_0 = (1; 4)(r_{01}; r_{02}) = r_{01} + 4r_{02} = 0, \quad |\mathbf{r}_0| = \sqrt{r_{01}^2 + r_{02}^2} = 1.$$

Поскольку точка $\mathbf{x}_0$ располагается только на одной (первой) граничной прямой ($i = 1$): $x_1 + 4x_2 = 14$, то условие (7.38) записывается в форме равенства.

Система ограничительных уравнений этой задачи имеет только два решения: $(-0,9700; 0,2425)$ и $(0,9700; -0,2425)$. Непосредственной подстановкой их в функцию T_0 устанавливаем, что максимум T_0 отличен от нуля и достигается при решении $(-0,9700; 0,2425)$. Таким образом, перемещаться из $\mathbf{x}_0$ нужно по направлению вектора $\mathbf{r}_0 = (-0,9700; 0,2425)$, т. е. по граничной прямой BA.

Для определения координат следующей точки $\mathbf{x}_1 = (x_{11}; x_{12})$:

$$\left.\begin{aligned} x_{11} &= x_{01} + r_{01}\lambda_0 = 4 - 0,9700\lambda_0, \\ x_{12} &= x_{02} + r_{02}\lambda_0 = 2,5 + 0,2425\lambda_0 \end{aligned}\right\} \tag{7.42}$$

необходимо найти значение λ_0^* параметра λ_0, при котором функция $f(\mathbf{x})$ в точке $\mathbf{x}_1$ достигает возможно бо́льшего значения. Но сначала найдем интервал допустимых значений параметра λ_0, при которых точка $\mathbf{x}_1$ будет принадлежать допустимой области. Система (7.36) в данном случае имеет вид

$$\left.\begin{aligned} 7(4 - 0,9700\lambda_0) + 3(2,5 + 0,2425\lambda_0) &\leqslant 42, \\ 4 - 0,9700\lambda_0 &\geqslant 0, \\ 2,5 + 0,2425\lambda_0 &\geqslant 0. \end{aligned}\right\} \tag{7.43}$$

Первое ограничение мы опустили, поскольку точка $\mathbf{x}_1$ лежит на прямой AB. Решая систему (7.43), устанавливаем, что λ_0 следует искать в полуинтервале (0; 4,1237] (заметим, что нас интересуют только неотрицательные значения параметра λ_0!). Найдем значение λ_0^*, при котором достигается наибольшее возрастание приращения Δf функции, вызванное перемещением из точки $\mathbf{x}_0$ в точку $\mathbf{x}_1$. В соответствии с условием (7.41)

$$\begin{gathered} d(\Delta f)/d\lambda_0 = \nabla f(\mathbf{x}_1) \cdot \mathbf{r}_0 = \\ = (-0,6 + 0,388\lambda_0;\ 1 - 0,0970\lambda_0)(-0,9700;\ 0,2425) = \\ = -0,3999\lambda_0 + 0,8245 = 0, \end{gathered}$$

откуда $\lambda_0 = 2,0618$. При этом $d^2(\Delta f)/d\lambda_0^2 = -0,3999 < 0$. Значит, $\lambda_0^* = 2,0618$. По формуле (7.42) находим координаты новой точки: $\mathbf{x}_1 = (2; 3)$.

Если продолжить оптимизационный поиск, то при решении очередной вспомогательной задачи (7.37) — (7.40) будет установлено, что $T_1 = \nabla f(\mathbf{x}_1) \cdot \mathbf{r}_1 = 0$, а это говорит о том, что точка $\mathbf{x}_1$ является точкой максимума $\mathbf{x}^*$ целевой функции в допустимой области. Это же видно и из рис. 7.8: в точке $\mathbf{x}_1$ одна из линий уровня касается границы допустимой области. Следовательно, точка $\mathbf{x}_1$ является точкой максимума $\mathbf{x}^*$. При этом $f_{\max} = f(\mathbf{x}^*) = 5,4$.

Задача с нелинейными ограничениями. Если в задачах с линейными ограничениями движение по граничным прямым оказывается возможным и даже целесообразным, то при нелинейных ограничениях, определяющих выпуклую область, любое сколь угодно малое перемещение из граничной точки может сразу вывести за пределы области допустимых решений, и возникнет необходимость в возвращении в допустимую область (рис. 7.9). Подобная ситуация характерна для задач, в которых экстремум функции $f(\mathbf{x})$ достигается на границе области. В связи с этим применяются различные способы перемещения, обеспечивающие построение последовательности точек, расположенных вблизи границы и внутри допустимой области, или зигзагообразное движение вдоль границы с пересечением последней. Как видно из рисунка, возврат из точки $\mathbf{x}_1$ в допустимую область следует осуществлять вдоль градиента той граничной функции $\varphi_k(\mathbf{x})$, которая оказалась

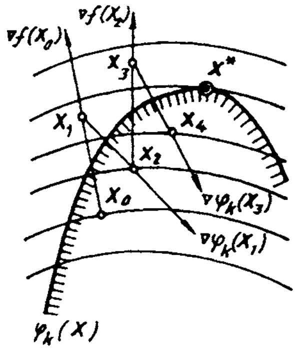


Рис. 7.9

нарушенной. Это обеспечит отклонение очередной точки $\mathbf{x}_2$ в сторону точки экстремума $\mathbf{x}^*$. Признаком экстремума в подобном случае будет коллинеарность векторов ∇f и $\nabla\varphi_k$.

7.5. ТЕОРЕМА КУНА – ТАККЕРА

В теории нелинейного программирования центральное место занимает теорема Куна — Таккера, обобщающая классический метод множителей Лагранжа на случай, когда в нелинейной задаче, помимо ограничений-равенств, содержатся также и неравенства. В частности, для задачи выпуклого программирования: минимизировать $z = f(\mathbf{x})$ при ограничениях $\varphi_i(\mathbf{x}) \leqslant 0 \ \ (i = \overline{1,m}\,)$, $\mathbf{x} \geqslant \mathbf{0}$, где все функции f и φ_i выпуклые, теорема Куна — Таккера устанавливает связь между оптимальным решением задачи и седловой точкой функции Лагранжа для этой задачи:

$$L(\mathbf{x}, \overline{\lambda}\,) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\mathbf{x}). \qquad (7.44)$$

Точка $(\mathbf{x}^*, \overline{\lambda}^*)$ называется *седловой точкой функции* (7.44), если n-мерная точка $\mathbf{x}^*$ является точкой минимума функции $L(\mathbf{x}, \overline{\lambda}^*)$, а m-мерная точка $\overline{\lambda}^*$ — точкой максимума функции $L(\mathbf{x}^*, \overline{\lambda})$, так что для всех $\mathbf{x}$ и $\overline{\lambda}$ выполняется неравенство

$$L(\mathbf{x}^*, \overline{\lambda}) \leqslant L(\mathbf{x}^*, \overline{\lambda}^*) \leqslant L(\mathbf{x}, \overline{\lambda}^*). \tag{7.45}$$

В этом смысл теоремы Куна — Таккера. Сама же теорема формулируется следующим образом.

***Теорема 7.7* (Куна — Таккера).** *Предположим, что существует вектор* $\mathbf{x} \geqslant \mathbf{0}$, *такой, что* $\varphi_i(\mathbf{x}) < 0 \quad (i = \overline{1, m})$. *Тогда необходимым и достаточным условием оптимальности вектора* $\mathbf{x}^*$, *принадлежащего допустимой области, является существование такого вектора* $\overline{\lambda}^*$, *что для всех* $\mathbf{x} \geqslant \mathbf{0}$ *и* $\overline{\lambda} \geqslant \mathbf{0}$ *имеет место неравенство* (7.45).

Первоначально эта теорема была доказана только для случая дифференцируемых функций $f(\mathbf{x})$ и $\varphi_i(\mathbf{x})$. Обобщение на случай выпуклых функций принадлежит Слейтеру.

Если функции $f(\mathbf{x})$ и $\varphi_i(\mathbf{x})$ являются дифференцируемыми, то неравенства (7.45), где $\mathbf{x}^* \geqslant \mathbf{0}$, $\overline{\lambda}^* \geqslant \mathbf{0}$, эквивалентны следующим "локальным" условиям Куна — Таккера:

$$\partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \mathbf{x} \geqslant 0, \tag{7.46}$$

$$\mathbf{x}^* \partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \mathbf{x} = 0, \tag{7.47}$$

$$\mathbf{x}^* \geqslant \mathbf{0}, \tag{7.48}$$

$$\partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \overline{\lambda} \leqslant 0, \tag{7.49}$$

$$\overline{\lambda}^* \partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \overline{\lambda} = 0, \tag{7.50}$$

$$\overline{\lambda}^* \geqslant \mathbf{0}. \tag{7.51}$$

Покажем, например, что если выполняются условия (7.46) — (7.51), то выполняются и условия (7.45), где $\mathbf{x}^* \geqslant \mathbf{0}$, $\overline{\lambda}^* \geqslant \mathbf{0}$. При фиксированном $\overline{\lambda}^*$ функция Лагранжа (7.44) $L(\mathbf{x}, \overline{\lambda}^*)$ является выпуклой по $\mathbf{x}$, а поэтому в соответствии с соотношением (7.11) имеет место неравенство $L(\mathbf{x}, \overline{\lambda}^*) - L(\mathbf{x}^*, \overline{\lambda}^*) \geqslant \partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \mathbf{x} \cdot (\mathbf{x} - \mathbf{x}^*)$, откуда с учетом равенства (7.47) получаем $L(\mathbf{x}, \overline{\lambda}^*) \geqslant L(\mathbf{x}^*, \overline{\lambda}^*) + \mathbf{x} \cdot \partial L(\mathbf{x}^*, \overline{\lambda}^*)/\partial \mathbf{x}$.

Поскольку $\mathbf{x} \geqslant \mathbf{0}$, то, опуская неотрицательное второе слагаемое в правой части, получаем лишь усиленное неравенство $L(\mathbf{x}, \overline{\lambda}^*) \geqslant L(\mathbf{x}^*, \overline{\lambda}^*)$. Таким образом, доказана выполнимость правого неравенства в соотношении (7.45). Аналогично показывается выполнимость для $\overline{\lambda} \geqslant \mathbf{0}$ и левого неравенства в соотношении (7.45).

Так же нетрудно устанавливается, что из условий (7.45), где $\mathbf{x}^* \geqslant \mathbf{0}$, $\overline{\lambda}^* \geqslant \mathbf{0}$, следуют условия (7.46) – (7.51).

7.6. ЗАДАЧА КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ И ЕЕ РЕШЕНИЕ

Одним из частных видов задачи выпуклого программирования является задача, в которой целевая функция содержит квадратичное слагаемое, а ограничения носят линейный характер. В этом случае говорят, что задача относится к *квадратичному программированию.*

В качестве основной в квадратичном программировании рассматривается задача минимизации функции

$$f(\mathbf{x}) = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_i x_j \tag{7.52}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i \quad (i = \overline{1, n}), \tag{7.53}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}). \tag{7.54}$$

Матрица $D = [d_{ij}]_{n \times n}$ квадратичной формы $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} d_{ij} x_i x_j$ предполагается симметрической и неотрицательно определенной. В этом случае функция (7.52) будет выпуклой.

Если в задаче квадратичного программирования целевая функция не минимизируется, а максимизируется или если в некоторых ограничениях вместо знака $\leqslant$ стоит знак $\geqslant$, то такие задачи всегда можно привести к основной форме (7.52) – (7.54).

С геометрической точки зрения задача (7.52) — (7.54) сводится к определению точки выпуклого многогранника решений, через которую проходит линия уровня поверхности $f(\mathbf{x})$, имеющая наименьшее значение функции $f(\mathbf{x})$. Если в разрешимой задаче линейного программирования оптимальное решение находилось в вершине многогранника, то целевая функция задачи (7.52) — (7.54) свой конечный минимум, даже если он единственный, может принимать как на границе, так и внутри многогранника решений.

Переходя к решению задачи (7.52) — (7.54), составим локальные условия Куна — Таккера (7.46) — (7.51), являющиеся необходимыми и достаточными условиями оптимальности решения $\mathbf{x}^*$. Функция Лагранжа в данном случае имеет вид

$$L = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{n}\sum_{j=1}^{n} d_{kj} x_k x_j + \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right). \quad (7.55)$$

Найдем частные производные функции (7.55):

$$\partial L/\partial x_j = c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} \quad (j = \overline{1,n}), \quad (7.56)$$

$$dL/d\lambda_i = \sum_{j=1}^{n} a_{ij} x_j - b_i \quad (i = \overline{1,m}). \quad (7.57)$$

В равенствах (7.56) и (7.57) введем обозначения:

$$c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j \quad (j = \overline{1,n}),$$

$$\sum_{j=1}^{n} a_{ij} x_j - b_i = -y_i \quad (i = \overline{1,m}).$$

С учетом этих обозначений условия (7.46) — (7.51), записанные в координатной форме, примут следующий вид:

$$\left.\begin{aligned} & c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j \geqslant 0, \\ & x_j v_j = 0, \quad x_j \geqslant 0 \quad (j = \overline{1,n}), \end{aligned}\right\} \quad (7.58)$$

$$\left.\begin{aligned} &\sum_{j=1}^{n} a_{ij}x_j - b_i = -y_i \leqslant 0, \\ &\lambda_i(-y_i) = 0, \quad \lambda_i \geqslant 0 \quad (i = \overline{1,m}). \end{aligned}\right\} \tag{7.59}$$

Объединяя соотношения (7.58) и (7.59), окончательно получаем:

$$\sum_{j=1}^{n} a_{ij}x_j + y_i = b_i \quad (i = \overline{1,m}), \tag{7.60}$$

$$2\sum_{k=1}^{n} d_{kj}x_k - v_j + \sum_{i=1}^{m} \lambda_i a_{ij} = -c_j \quad (j = \overline{1,n}), \tag{7.61}$$

$$x_j \geqslant 0,\ v_j \geqslant 0\ (j = \overline{1,n}); \quad y_i \geqslant 0,\ \lambda_i \geqslant 0\ (i = \overline{1,m}), \tag{7.62}$$

$$\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0. \tag{7.63}$$

Равенства (7.60), (7.61) образуют систему $N = n + m$ линейных уравнений с $2N = 2(n + m)$ неизвестными $x_1, \ldots, x_n$, $v_1, \ldots, v_n$, $y_1, \ldots, y_m$, $\lambda_1, \ldots, \lambda_m$.

Итак, в соответствии с локальными условиями Куна — Таккера решение $\mathbf{x}^* = (x_1^*; \ldots; x_n^*)$ является оптимальным для задачи (7.52) — (7.54) тогда и только тогда, когда совместно с решением $\mathbf{v} = (v_1; \ldots; v_n)$ существуют решения $\overline{\lambda} = (\lambda_1; \ldots; \lambda_m)$, $\mathbf{y} = (y_1; \ldots; y_m)$, такие, что $\mathbf{z} = (x_1; \ldots; x_n; v_1; \ldots; v_n; \lambda_1; \ldots; \lambda_m; y_1; \ldots; y_m)$ является решением системы (7.60) — (7.62) при условии выполнения равенства (7.63).

Таким образом, применение теоремы Куна — Таккера для решения задачи квадратичного программирования позволяет нелинейную задачу сводить к решению системы линейных алгебраических уравнений при соблюдении, правда, дополнительного комбинаторного условия (7.63). Это условие требует, чтобы из каждых двух ограниченных по знаку переменных x_j и v_j (соответственно y_i и λ_i) хотя бы одна равнялась нулю. Иначе говоря, не все решения $\mathbf{z} = (x_1; \ldots; y_m)$ системы (7.60), (7.61) удовлетворяют условию (7.63), а только те, в которых по крайней мере $n + m$ компонент равны нулю, т. е. столько, сколько уравнений в системе. Но таким свойством обладают базисные решения системы. Значит, искать решение, которое удовлетворяло бы условию (7.63), имеет смысл только среди базисных решений. Известно, что с базисными решениями оперирует симплексный метод. Воспользуемся им, внеся в вычислительную процедуру определенные изменения, обусловленные спецификой рассматриваемой задачи.

Запишем равенства (7.60) — (7.63) в векторно-матричной форме

$$A\mathbf{x} + E\mathbf{y} + O\mathbf{v} + O\overline{\lambda} = \mathbf{b}, \tag{7.64}$$

$$2D\mathbf{x} + O\mathbf{y} - E\mathbf{v} + A^T\overline{\lambda} = -\mathbf{c}, \tag{7.65}$$

$$\mathbf{x} \geqslant \mathbf{0},\ \mathbf{y} \geqslant \mathbf{0},\ \mathbf{v} \geqslant \mathbf{0},\ \overline{\lambda} \geqslant \mathbf{0}, \tag{7.66}$$

$$\mathbf{x}' \cdot \mathbf{v} + \mathbf{y}' \cdot \overline{\lambda} = 0, \tag{7.67}$$

где

$$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix},\ E = \begin{bmatrix} 1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & 1 \end{bmatrix},\ D = \begin{bmatrix} d_{11} & \ldots & d_{1n} \\ \ldots & \ldots & \ldots \\ d_{n1} & \ldots & d_{nn} \end{bmatrix},$$

$$A^T = \begin{bmatrix} a_{11} & \ldots & a_{m1} \\ \ldots & \ldots & \ldots \\ a_{1n} & \ldots & a_{mn} \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} b_1 \\ \ldots \\ b_m \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} c_1 \\ \ldots \\ c_n \end{bmatrix}.$$

Объединим равенства (7.64) и (7.65) следующей записью:

$$\begin{bmatrix} A & E & O & O \\ 2D & O & -E & A^T \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{v} \\ \overline{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{c} \end{bmatrix}. \tag{7.68}$$

В целях дальнейшего упрощения записей обозначим в равенстве (7.68) через $\mathbf{z}$ вектор $\mathbf{z}' = (\mathbf{x}', \mathbf{y}', \mathbf{v}', \overline{\lambda'})$ и одновременно будем рассматривать вектор $\widetilde{\mathbf{z}'} = (\mathbf{v}', \overline{\lambda'}, \mathbf{x}', \mathbf{y}')$. Ясно, что компоненты векторов $\mathbf{z}$ и $\widetilde{\mathbf{z}}$ попарно равны, т. е.

$$\widetilde{z}_i = z_{i+N}, \quad \widetilde{z}_{i+N} = z_i \quad (i = \overline{1, N}). \tag{7.69}$$

Ввиду того что

$$\mathbf{z} \cdot \widetilde{\mathbf{z}} = \mathbf{x}' \cdot \mathbf{v} + \mathbf{y}' \cdot \overline{\lambda} + \mathbf{v}' \cdot \mathbf{x} + \overline{\lambda'} \cdot \mathbf{y} = 2(\mathbf{x}' \cdot \mathbf{v} + \mathbf{y}' \cdot \overline{\lambda}),$$

равенство (7.67) можно записать в виде $T = \mathbf{z}' \cdot \widetilde{\mathbf{z}} = 0$, а условие (7.66) — в виде $\mathbf{z} \geqslant \mathbf{0}$. Учитывая равенство (7.68) и другие введенные обозначения, окончательно локальные условия Куна — Таккера (7.60) — (7.63) запишем так:

$$\begin{bmatrix} A & E & O & O \\ 2D & O & -E & A^T \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{c} \end{bmatrix}, \tag{7.70}$$

$$\mathbf{z} \geqslant \mathbf{0}, \tag{7.71}$$

$$T = \mathbf{z}' \cdot \widetilde{\mathbf{z}} = 0. \tag{7.72}$$

Задача состоит в том, чтобы среди базисных решений системы (7.70), (7.71) найти такое решение $\mathbf{z}^*$, которое обращает в нуль выпуклую функцию T, заданную выражением (7.72).

Последовательный перебор базисных решений осуществим с помощью симплексных преобразований. Перебор начинается с некоторого исходного базисного решения $\mathbf{z}_0$ и организуется так, чтобы при переходе от одного решения к другому значения функции T убывали. Симплексные преобразования выполняют по тем же правилам, что и в линейном программировании, но свободная переменная, вводимая в базис, выбирается по специальному правилу, к выводу которого мы и переходим.

Предположим, что в системе (7.70) все $N = n + m$ уравнений линейно независимы и система разрешена, например, относительно первых N неизвестных, т. е. имеет вид

$$z_k = h_{k0} - \sum_{t=1}^{N} h_{kt} z_{N+1} \quad (k = \overline{1, N}). \tag{7.73}$$

Такой записи системы соответствует симплексная таблица (табл. 7.1).

Таблица 7.1

	1	$-z_{N+1}$	...	$-z_{2N}$
$z_1 =$	h_{10}	h_{11}	...	h_{1N}
...	...	....	....	....
$z_N =$	h_{N0}	h_{N1}	...	h_{NN}

Равенство типа (7.73) можно записать и для каждой свободной переменной. Оно будет иметь вид $z_{N+s} = 0 + 1 \cdot z_{N+s}$. Таким равенствам в симплексной таблице будут отвечать строки, все элементы которых, кроме одного (равного -1), равны нулю (табл. 7.2). При таком соглашении в левый заглавный столбец будут входить все переменные z_k $(k = \overline{1, 2N})$ и их можно расположить в порядке возрастания индексов.

Таблица 7.2

	1	$-z_{N+1}$	$\ldots$	$-z_{2N}$
$z_1 =$	h_{10}	h_{11}	$\ldots$	h_{1N}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$z_N =$	h_{N0}	h_{N1}	$\ldots$	h_{NN}
$z_{N+1} =$	0	-1	$\ldots$	0
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$z_{2N} =$	0	0	$\ldots$	-1

Заменим координатные равенства (7.73) одним — векторным:

$$\mathbf{z} = \mathbf{h}_0 - \sum_{t=1}^{N} \mathbf{h}_t z_{N+t}, \tag{7.74}$$

где $\mathbf{h}_t$ — t-й вектор-столбец табл. 7.2.

При нулевых значениях свободных переменных z_{N+t} из равенства (7.74) находим начальное базисное решение:

$$\mathbf{z}_0 = \mathbf{h}_0 \geqslant \mathbf{0}. \tag{7.75}$$

Вектор $\widetilde{\mathbf{z}}_0$, соответствующий вектору (7.75), обозначим через $\widetilde{\mathbf{h}}_0$. Тогда значение функции (7.72), соответствующее начальному базисному решению (7.75), будет равно

$$T_0 = T(\mathbf{h}_0) = \mathbf{h}_0' \cdot \widetilde{\mathbf{h}}_0. \tag{7.76}$$

Выполним симплексное преобразование, выбрав разрешающий элемент по правилам, известным из линейного программирования. Найденное при этом минимальное симплексное отношение обозначим через θ_s, т. е.

$$\theta_s = \min_{h_{ks}>0} (h_{k0}/h_{ks}).$$

В результате симплексного преобразования получим новое базисное решение $\mathbf{z}_1$, которое в соответствии с правилами пересчета элементов симплексных таблиц будет иметь вид

$\mathbf{z}_1 = \mathbf{h}_0 - \theta_s \mathbf{h}_s$, а соответствующий ему вектор $\widetilde{\mathbf{z}}_1$ примет вид $\widetilde{\mathbf{h}}_0 - \theta_s \widetilde{\mathbf{h}}_s$. Найдем значение функции $T(\mathbf{z})$, соответствующее новому базисному решению:

$$T_s = T(\mathbf{z}_1) = \mathbf{z}_1' \cdot \widetilde{\mathbf{z}}_1 = (\mathbf{h}_0 - \theta_s \mathbf{h}_s)(\widetilde{\mathbf{h}}_0 - \theta_s \widetilde{\mathbf{h}}_s) =$$

$$= \mathbf{h}_0' \cdot \widetilde{\mathbf{h}}_0 - \theta_s \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_0 - \mathbf{h}_0' \theta_s \widetilde{\mathbf{h}}_s + \theta_s^2 \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_s.$$

Из равенства (7.69) следует, что $\mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_0 = \mathbf{h}_0' \cdot \widetilde{\mathbf{h}}_s$, поэтому $T_s = \mathbf{h}_0' \cdot \widetilde{\mathbf{h}}_0 - 2\theta_s \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_0 + \theta_s^2 \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_s$. Учитывая равенство (7.76), получаем

$$T_s = T_0 + \theta_s(-2\mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_0 + \theta_s \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_s). \tag{7.77}$$

Обозначим в уравнении (7.77) $\alpha_s = \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_0$, $\beta_s = \mathbf{h}_s' \cdot \widetilde{\mathbf{h}}_s$. Тогда

$$T_s = T_0 + \theta_s(-2\alpha_s + \theta_s \beta_s). \tag{7.78}$$

Обозначая в уравнении (7.78)

$$K_s = -2\alpha_s + \theta_s \beta_s, \tag{7.79}$$

приходим к равенству

$$T_s = T_0 + \theta_s K_s. \tag{7.80}$$

Так как значение функции T при симплексных преобразованиях должно уменьшаться, в базис надо вводить ту свободную переменную z_{N+s}, для которой $K_s < 0$, ибо, как видно из равенства (7.80), только в этом случае разность $T_s - T_0 < 0$, т. е. $T_s < T_0$. Напомним, что величина θ_s всегда положительна. Величины β_s, являющиеся по существу вторыми производными от T по θ_s, всегда неотрицательны, поэтому величина K_s может быть отрицательной, как это видно из равенства (7.79), только когда $\alpha_s > 0$.

Суть нового правила выбора свободной переменной, включаемой в базис, состоит в том, что каждую свободную переменную z_{N+s}, для которой $\alpha_s > 0$, проверяют: уменьшит ли базисное решение, полученное после включения этой переменной в базис, значение функции T.

Может случиться, что $K_s > 0$ для всех s, хотя все еще $T > 0$. Тогда можно начать процесс оптимизационного поиска сначала, взяв другое начальное базисное решение, либо же принять временное увеличение функции T, включая в базис свободную переменную z_{N+s} с положительным K_s и надеясь при следующих итерациях понизить значение T. В этом состоит недостаток рассматриваемого метода.

Существуют и другие методы решения задачи (7.52) — (7.54), но и они не свободны от недостатков.

Подытоживая рассуждения, можно предложить следующий порядок решения задачи (7.52) — (7.54) рассмотренным способом:

1) записать задачу в форме (7.52) — (7.54);

2) по данным задачи составить локальные условия Куна — Таккера в форме (7.70) — (7.72);

3) найти начальное базисное решение $\mathbf{h}_0$;

4) на основе найденного базисного решения минимизировать функцию $T = \mathbf{z}'\widetilde{\mathbf{z}}$. С этой целью составить таблицу, включив в ее основную часть строки для всех переменных, расположив их в порядке возрастания индекса k переменной z_k, т. е. в порядке $x_1, \ldots, x_n$, $y_1, \ldots, y_m$, $v_1, \ldots, v_n$, $\lambda_1, \ldots, \lambda_m$ (табл. 7.3).

Таблица 7.3

<table>
<tr><td></td><td></td><td>1</td><td>$-z_{N+1} \ldots - z_{2N}$</td></tr>
<tr><td rowspan="3">Основная часть</td><td>$(z_1 =)\ x_1 =$</td><td>h_{10}</td><td>$h_{11} \ldots h_{1N}$</td></tr>
<tr><td>.............</td><td>...</td><td>...............</td></tr>
<tr><td>$(z_{2N} =)\ \lambda_m =$</td><td>$h_{2N,0}$</td><td>$h_{2N,1} \ldots h_{2N,N}$</td></tr>
<tr><td rowspan="5">Дополнительная часть</td><td>T</td><td></td><td></td></tr>
<tr><td>α_s</td><td></td><td>$\alpha_1 \ldots \alpha_N$</td></tr>
<tr><td>β_s</td><td></td><td>$\beta_1 \ldots \beta_N$</td></tr>
<tr><td>θ_s</td><td></td><td>$\theta_1 \ldots \theta_N$</td></tr>
<tr><td>K_s</td><td></td><td>$K_1 \ldots K_N$</td></tr>
</table>

Дополнительную часть таблицы заполнять в следующем порядке:

а) вычислить $T_0 = \mathbf{h}_0'\widetilde{\mathbf{h}}_0$, где $\mathbf{z}_0 = \mathbf{h}_0$ — начальное базисное решение. Если $T_0 = 0$, то данная задача решена и останется составить оптимальное решение $\mathbf{x}^*$ из компонент вектора $\mathbf{h}_0$ и вычислить $f_{\min}$. Если же $T_0 \neq 0$, то вычисления продолжить;

б) найти для всех $s = \overline{1, N}$ $\alpha_s = \mathbf{h}'_s \widetilde{\mathbf{h}}_0$;

в) для тех s, для которых $\alpha_s > 0$, вычислить $\beta_s = \mathbf{h}'_s \widetilde{\mathbf{h}}_s$, $\theta_s = \min h_{k0}/h_{ks}$ (элементы h_{ks}, для которых отношение θ_s минимально, отметить в табл. 7.3); вычислить $K_s = -2\alpha_s + \theta_s \beta_s$;

г) выбрать разрешающий элемент для выполнения симплексного преобразования. При этом разрешающий столбец определить по отрицательному элементу K_s с наибольшей абсолютной величиной. Элемент h_{ks} выбранного столбца, которому соответствует наименьшее отношение θ_s, становится разрешающим. С ним выполнить симплексное преобразование и получить новое базисное решение.

Операции п. 4 алгоритма выполнять до тех пор, пока функция T не примет нулевое значение. Если все $K_s > 0$ и $T > 0$, то включить в базис переменную, отвечающую положительному K_s, или выбрать в качестве начального другое базисное решение.

7.7. МЕТОДЫ ШТРАФНЫХ И БАРЬЕРНЫХ ФУНКЦИЙ

Методы решения задач нелинейного, в частности выпуклого, программирования, при использовании которых данную задачу можно свести к задаче минимизации некоторой специальной функции, представляющей собой сумму данной минимизируемой функции и некоторой другой функции (называемой *штрафной*), сформированной из ограничительных функций данной задачи, называют *методами штрафных функций*. Идея этих методов состоит в замене целевой функции данной задачи некоторой обобщенной функцией, значения которой совпадают со значениями исходной функции внутри допустимой области, но при приближении к границе области, а тем более при выходе из нее, резко возрастают за счет второго слагаемого обобщенной функции — штрафной функции. Штрафные функции строятся таким образом, что обеспечивают либо быстрое возвращение в допустимую область, либо невозможность выхода из нее. Методы штрафных функций сводят задачу на условный экстремум к решению последовательности задач на безусловный экстремум, что нередко оказывается значительно проще. Эффективность такого подхода становится особенно ощутимой, когда ограничения исходной задачи заданы нелинейными функциями. В зависимости от способа формирования штрафных функций различают метод штрафных и метод барьерных функций.

Рассмотрим метод штрафных функций. Пусть требуется минимизировать функцию

$$z = f(\mathbf{x}) \tag{7.81}$$

при ограничениях

$$\varphi_i(\mathbf{x}) \leqslant 0 \quad (i = \overline{1, m}). \tag{7.82}$$

Предполагаем, что ограничения $\mathbf{x} \geqslant \mathbf{0}$ включены в ограничения (7.82). Функция $T(\mathbf{x}, t)$, обобщенная для функции (7.81), имеет вид

$$T(\mathbf{x}, t) = f(\mathbf{x}) + t\theta(\mathbf{x}), \tag{7.83}$$

где t — некоторое положительное число, называемое *коэффициентом штрафа*; $\theta(\mathbf{x})$ — непрерывная функция штрафа, удовлетворяющая условиям: $\theta(\mathbf{x}) = 0$ для всех точек $\mathbf{x}$ допустимой области и $\theta(\mathbf{x}) > 0$ для всех остальных точек.

Хорошо изучены штрафные функции видов:

$$\theta_1(\mathbf{x}) = \sum_{i=1}^{m} (\max\{\varphi_i(\mathbf{x}), 0\})^\alpha$$

(где $\alpha = 1$; $\alpha = 2$) и

$$\theta_2(\mathbf{x}) = \sum_{i=1}^{m} 0,25(\varphi_i(\mathbf{x}) + |\varphi_i(\mathbf{x})|)^2.$$

Процедура оптимизационного поиска по методу штрафных функций состоит в следующем. Рассматривается некоторая неограниченная монотонно возрастающая последовательность $\{t_k\}$ $(k = 1, 2, \dots)$ положительных чисел. Для первого числа t_1 этой последовательности находится точка $\mathbf{x}_1^*$, доставляющая минимум функции (7.83). Найденная точка $\mathbf{x}_1^*$ используется как начальное приближение для решения задачи поиска минимума функции $T(\mathbf{x}, t_2)$, где $t_2 > t_1$ и т.д. Таким образом решается последовательность задач минимизации функции $T(\mathbf{x}, t_k)$ $(k = 1, 2, \dots)$, причем результат предыдущей оптимизации $\mathbf{x}_k^*$ используется в качестве начального приближения для поиска $\mathbf{x}_{k+1}^*$. Поскольку для бесконечно возрастающей последовательности $\{t_k\}$ локальные минимумы приближаются к допустимой области, то последовательность $\{\mathbf{x}_k^*\}$ $(k = 1, 2, \dots)$ сходится к локальному оптимуму функции $f(\mathbf{x})$, расположенному внутри или на границе допустимой области. Точки $\mathbf{x}_k^*$ расположены вне допустимой области, поэтому метод штрафных функций называют также *методом внешней точки.* В этом методе любая точка может быть выбрана в качестве начальной, что значительно упрощает машинное программирование алгоритма решения задачи.

Рассмотрим теперь *метод барьерных функций.* Если ограничения (7.82) имеют вид строгих неравенств, то для формирования обобщенной функции используются так называемые *барьерные функции* $I(\mathbf{x})$, значения которых неограниченно возрастают при приближении к границе допустимой области. Обобщенная функция $U(\mathbf{x}, r)$ имеет вид

$$U(\mathbf{x}, r) = f(\mathbf{x}) + rI(\mathbf{x}), \tag{7.84}$$

где r — некоторое положительное число.

Барьерная функция $I(\mathbf{x})$ должна быть непрерывной во всех точках, лежащих внутри допустимой области, и если $\{\mathbf{x}_k\}$ $(k = 1, 2, \dots)$ — последовательность внутренних точек, сходящихся к граничной точке допустимой области, то последовательность $\{I(\mathbf{x}_k)\}$ значений барьерной функции неограниченно возрастает. Поскольку все точки последовательности лежат в допустимой области, метод барьерных функций называют также *методом внутренней точки.*

В качестве барьерной часто применяется логарифмическая функция вида

$$I_1(\mathbf{x}) = -\sum_{i=1}^{m} \ln(-\varphi_i(\mathbf{x})),$$

определенная только внутри допустимой области. Она неограниченно возрастает при приближении к границе допустимой области.

Используется также более простая функция

$$I_2(\mathbf{x}) = -\sum_{i=1}^{m}(\varphi_i(\mathbf{x}))^{-1},$$

называемая обратной и определенная всюду, за исключением границы допустимой области.

Штрафная добавка $rI(\mathbf{x})$ к целевой функции $f(\mathbf{x})$ в равенстве (7.84) образует как бы барьер, препятствующий выходу из допустимой области.

В алгоритме оптимизационного поиска используется последовательность $\{r_k\}$ положительных чисел r_k $(k = 1, 2, \ldots)$, монотонно сходящаяся к нулю. В качестве начальной точки берут произвольную внутреннюю точку $\mathbf{x}_0$ допустимой области. Она является исходной для поиска точки минимума $\mathbf{x}_1^*$ обобщенной функции $U(\mathbf{x}, r_1)$. Точка $\mathbf{x}_1^*$ используется в качестве начального приближения для поиска точки $\mathbf{x}_2^*$ минимума функции $U(\mathbf{x}, r_2)$ и т. д. Последовательность $\{\mathbf{x}_k^*\}$ полученных таким образом точек безусловных минимумов функций $U(\mathbf{x}, r_k)$ сходится к точке минимума $\mathbf{x}^*$ функции $f(\mathbf{x})$ задачи. Приближение точек $\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots$ к оптимальной точке $\mathbf{x}^*$ осуществляется внутри допустимой области. При уменьшении r убывает влияние штрафной добавки $rI(\mathbf{x})$ в равенстве (7.84) и возрастает влияние целевой функции $f(\mathbf{x})$ задачи. Поэтому последовательность функций $U(\mathbf{x}, r_k)$ дает сколь угодно точное (для достаточно большого номера k) приближение к локальному минимуму функции $f(\mathbf{x})$. Если искомый экстремум лежит внутри допустимой области, то решение может быть получено после нескольких первых значений параметра r.

Трудности в расчетах возрастают при определении начальной точки $\mathbf{x}_0$ внутри допустимой области. Поиск усложняется, если экстремум достигается на границе области. Поскольку значения обобщенной функции при приближении к границе области определяются главным образом величиной барьерной функции, экстремум не всегда может быть вычислен с заданной точностью.

7.8. ПОНЯТИЕ О МЕТОДЕ ЛИНЕЙНОЙ АППРОКСИМАЦИИ

Функция n переменных $f(x_1, \ldots, x_n)$ называется *сепарабельной*, если для нее справедливо представление

$$f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f_n(x_n) = \sum_{j=1}^{n} f_j(x_j).$$

Будем рассматривать задачу, в которой все функции сепарабельные, т. е. задачу вида

$$\min(\max) f = \sum_{j=1}^{n} f_j(x_j); \tag{7.85}$$

$$\sum_{j=1}^{n} g_{ij}(x_j) \; \{\leqslant, =, \geqslant\} \; b_i \quad (i = \overline{1, m}). \tag{7.86}$$

Приближенное решение задачи довольно просто: для всех функций f_j и g_{ij} строятся их кусочно-линейные аппроксимации и нелинейная задача заменяется приближенной задачей линейного программирования, которая решается симплекс-методом. При этом находится ее локальный экстремум, который будет приближенным локальным экстремумом данной задачи. Если задача (7.85), (7.86) является задачей выпуклого программирования, то найденный приближенный локальный экстремум будет одновременно и глобальным.

Замену нелинейной функции ее кусочно-линейной аппроксимацией рассмотрим на примере непрерывной функции $f(x)$ одной переменной, определенной на отрезке $[c_1; c_2]$. Разделим отрезок $[c_1; c_2]$ точками $x_0 \equiv c_1, x_1, x_2, \ldots, x_{r-1}, x_r \equiv c_2$ и вычислим соответствующие значения функции $f(x)$ в точках деления $f(x_0), f(x_1), \ldots, f(x_r)$. Точки $(x_k; f(x_k))$ на кривой $f(x)$ соединим отрезками прямых линий. Мы получим таким образом кусочно-линейную аппроксимацию — функцию $\widetilde{f}(x)$, которая приближает данную функцию $f(x)$ (рис. 7.10). При измельчении частичных отрезков $[x_k; x_{k+1}]$ приближение уточняется.

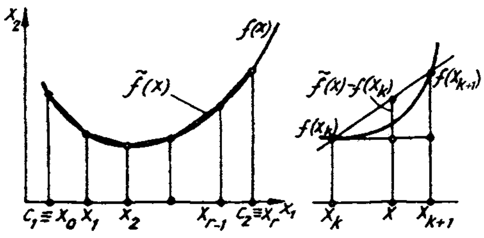


Рис. 7.10

Предположим, что в задаче (7.85), (7.86) все функции f_j и g_{ij} непрерывны. Разобьем отрезок, который может пробегать переменная x_j $(j = \overline{1,n})$, точками x_{kj} и, используя описанный способ, построим кусочно-линейные аппроксимации $\tilde f_j(x_j)$ и $\tilde g_{ij}(x_j)$ для всех функций $f_j(x_j)$ и $g_{ij}(x_j)$. Тем самым мы заменим задачу (7.85), (7.86) приближенной задачей:

$$\min(\max)\,\tilde f = \sum_{j=1}^{n} \tilde f_j(x_j); \tag{7.87}$$

$$\sum_{j=1}^{n} \tilde g_{ij}(x_j)\ \{\leqslant, =, \geqslant\}\ b_i \quad (i = \overline{1,m}). \tag{7.88}$$

Теперь определим локальный экстремум приближенной задачи (7.87), (7.88). С этой целью прежде всего найдем аналитический вид аппроксимирующей функции $\tilde f(x)$ на отрезке $[x_k; x_{k+1}]$. Пусть x — произвольная точка этого отрезка. Из рис. 7.10 видно, что

$$\tilde f(x) = f(x_k) + \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}(x - x_k). \tag{7.89}$$

Но каждый фиксированный x можно представить в виде $x = (1-\lambda)x_k + \lambda x_{k+1}$, где λ имеет вполне определенное значение, принадлежащее отрезку $[0;1]$. Подставляя $x - x_k = \lambda(x_{k+1} - x_k)$ в равенство (7.89), получаем

$$\tilde f(x) = (1-\lambda)f(x_k) + \lambda f(x_{k+1}). \tag{7.90}$$

Обозначив в равенстве (7.90) $1-\lambda=\lambda_k$ и $\lambda=\lambda_{k+1}$, будем иметь

$$\widetilde{f}(x)=\lambda_k f(x_k)+\lambda_{k+1}f(x_{k+1}), \tag{7.91}$$

где $\lambda_k \geqslant 0$; $\lambda_{k+1} \geqslant 0$; $\lambda_k+\lambda_{k+1}=1$;

$$x=\lambda_k x_k+\lambda_{k+1}x_{k+1}. \tag{7.92}$$

Обобщая выражения (7.91) и (7.92) таким образом, чтобы с их помощью можно было определить любое число $x \in [c_1; c_2]$ и любую соответствующую функцию $\widetilde{f}(x)$, получаем:

$$x=\sum_{k=0}^{r}\lambda_k x_k, \tag{7.93}$$

$$\widetilde{f}(x)=\sum_{k=0}^{r}\lambda_k f(x_k), \tag{7.94}$$

где все $\lambda_k > 0$ и $\sum\limits_{k=0}^{r}\lambda_k=1$. При этом следует потребовать, чтобы не более двух соседних λ_k были положительными, ибо только в этом случае точки, определяемые выражениями (7.93) и (7.94), будут лежать на аппроксимирующей ломаной.

Воспользовавшись рассмотренным способом аппроксимации, заменим в соотношениях (7.85), (7.86) все функции их кусочно-линейными приближениями:

$$\widetilde{f}_j(x_j)=\sum_{k=0}^{r_j}\lambda_{kj}f_j(x_{kj}), \tag{7.95}$$

$$\widetilde{g}_{ij}(x_j)=\sum_{k=0}^{r_j}\lambda_{kj}g_{ij}(x_{kj}), \tag{7.96}$$

где $x_j=\sum\limits_{k=0}^{r_j}\lambda_{kj}x_{kj}$; $\lambda_{kj} \geqslant 0$; $\sum\limits_{k=0}^{r_j}\lambda_{kj}=1$ при всех k,j $(i=\overline{1,m};\ j=\overline{1,n})$, причем для данного j не более двух соседних по индексу k значений λ_{kj} могут быть положительными.

Подчеркнем, что при построении функций $\widetilde{f}_j$ и $\widetilde{g}_{ij}$ используется одно и то же разбиение отрезка $[c_{1j}; c_{2j}]$. При этом точки x_{kj} выбираются так, чтобы обеспечить нужную точность аппроксимации всех функций f_j и g_{ij}.

Подставляя выражения (7.95) и (7.96) в соотношения (7.87) и (7.88), получаем приближенную задачу линейного программирования:

$$\min(\max)\,\widetilde{f} = \sum_{j=1}^{n}\sum_{k=0}^{r_j}\lambda_{kj}f_j(x_{kj}); \tag{7.97}$$

$$\sum_{j=1}^{n}\sum_{k=0}^{r_j}\lambda_{kj}g_{ij}(x_{kj})\ \{\leqslant, =, \geqslant\}\ b_i \quad (i=\overline{1,m}), \tag{7.98}$$

$$\sum_{k=0}^{r_j}\lambda_{kj} = 1 \quad (j=\overline{1,n}), \qquad \lambda_{kj}\geqslant 0 \quad (k=\overline{0,r_j}). \tag{7.99}$$

Заметим, что в выражениях (7.97), (7.98) $f_j(x_{kj})$ и $g_{ij}(x_{kj})$ — фиксированные числа, так что в приближенной задаче вместо переменных x_j используются переменные λ_{kj}.

Решая задачу (7.97) – (7.99) симплекс-методом при дополнительном условии, что в базис могут одновременно входить при каждом j не более двух соседних по индексу k переменных λ_{kj}, находят λ^*_{kj}, а затем по формуле

$$x_j^* = \sum_{k=0}^{r_j}\lambda^*_{kj}x_{kj}$$

определяют значения координат точки экстремума приближенной задачи.

Для упрощения вычислений переменные, входящие в задачу линейно, не следует выражать через λ_{kj}.

Заметим еще, что найденная точка локального экстремума, вообще говоря, не обязательно будет лежать в допустимой области исходной задачи. Однако нарушения ограничений, как правило, незначительны. Если же допустимая область выпуклая, то любое допустимое решение приближенной задачи находится в допустимой области исходной задачи.

8. ЭЛЕМЕНТЫ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

8.1. ПРИМЕРЫ ЗАДАЧ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ, ИХ ОСОБЕННОСТИ И ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ

Основные понятия. *Динамическое программирование* (иначе — *динамическое планирование*) — это метод нахождения оптимальных решений в задачах с многошаговой (многоэтапной) структурой. Многие экономические процессы расчленяются на шаги естественным образом. Это все процессы планирования и управления, развивающиеся во времени. Естественным шагом в них может быть год, квартал, месяц, декада, неделя, день и т. д. Однако метод динамического программирования может использоваться при решении задач, где время вообще не фигурирует; разделение на шаги в таких задачах вводится искусственно. Поэтому "динамика" задач динамического программирования заключается в методе решения.

В экономической практике встречается несколько типов задач, которые по постановке или способу решения относятся к задачам динамического программирования. Это задачи оптимального перспективного и текущего планирования во времени. Их решают либо путем составления комплекса взаимосвязанных статических моделей для каждого периода, либо путем составления единой динамической задачи оптимального программирования с применением многошаговой процедуры принятия решений. К задачам динамического программирования следует отнести задачи многошагового нахождения оптимума при размещении производительных сил, а также оптимального быстродействия. Смысл задач последнего типа состоит в следующем. Известна некоторая оптимальная структура производства с эффективностью Z_1. В начальный момент времени существует неоптимальная структура с эффективностью Z_0. Необходимо определить такие управляющие воздействия, которые за кратчайший период переведут структуру из начального положения в оптимальное (задача оптимальной перестройки).

Приведем еще несколько типичных задач, для решения которых естественным является применение метода динамического программирования.

Задача перспективного планирования. Планируется деятельность группы N промышленных предприятий Π_i $(i = \overline{1,N})$ на период в t $(t = \overline{1,T})$ хозяйственных лет. В начале периода на развитие системы предприятий выделены какие-то средства K, которые должны быть распределены между предприятиями. В процессе деятельности предприятия вложенные в него средства частично амортизируются. Каждое предприятие за год приносит доход, зависящий от вложенных средств, часть которого отчисляется в фонд предприятий. В начале каждого хозяйственного года имеющиеся средства перераспределяются между предприятиями. Возникает задача определения объема средств в начале каждого года, которые нужно выделить каждому предприятию, чтобы суммарный чистый доход за T лет был максимальным. Это типичная задача динамического программирования. Здесь процесс принятия решения разбивается на T шагов. Управление им заключается в начальном распределении и последующих перераспределениях средств $u^t = \{u_i^t\}$, где u_i^t — объем средств, выделенных i-му предприятию в начале t-го года. Для описания динамики системы вводится вектор состояния $\mathbf{x}^t = \{\mathbf{x}_i^t\}$, где $\mathbf{x}_i^t$ — состояние i-го предприятия на начало t-го года. В свою очередь состояние каждого предприятия $\mathbf{x}_i^t$ является вектором, компонентами которого служат трудовые ресурсы, основные фонды, финансовое положение и т.д., т.е. $\mathbf{x}_i^t = \{x_{ik}^t\}$. Очевидно, что вектор управления — это функция состояния системы на начало соответствующего года: $\mathbf{u}^t = \mathbf{u}^t(\mathbf{x}^{t-1})$. Начальное состояние системы $\mathbf{x}^0$ может быть заданным. Целевой функцией будет суммарная прибыль объединения за T лет. Если z^t — прибыль за t-й год, то получим задачу

$$\max Z = \sum_{t=1}^{T} z^t, \quad u \in \Omega,$$

где Ω — область допустимых управлений, или множество экономических возможностей, определяемых различными ограничениями, налагаемыми на состояние системы и вектор управления.

Задача об оптимальном управлении поставками. В различных областях народного хозяйства возникает задача определения момента подачи партии поставки и ее объема. С размещением заказов связаны некоторые фиксированные затраты K, не зависящие от величины заказываемой партии, а зависящие только от факта заказывания. С содержанием материальных ресурсов связаны затраты, пропорциональные остатку нереализованной продукции на конец интервала. Пусть T — промежуток планирования. Обозначим через ν_t интенсивность потребления ресурса в t-м интервале. Состояние системы будем описывать величиной остатка нереализованной продукции на конец интервала x_t. Начальное x_0 и конечное x_T состояния системы можно считать заданными. Для бесперебойности потребления поставками нужно управлять. Обозначим через $\mathbf{u} = \{u_t\}$ вектор управления, координаты которого суть величины поставок в начале соответствующих интервалов. Очевидно, что вектор управления есть функция состояния на начало интервала. Из множества возможных управлений требуется выбрать такое, при котором достигается минимум издержек на заказывание и содержание материальных ресурсов. Если S_t — издержки содержания единицы продукции в t-м интервале, то функция цели примет вид

$$\min Z = \sum_{t=1}^{T}(k\delta_t + S_t x_t),$$

где $\delta_t = \begin{cases} 1, & \text{если } u_t > 0, \\ 0, & \text{если } u_t = 0. \end{cases}$

Состояние системы опишется соотношением $x_t = x_{t-1} + u_t - \nu_t \quad (t = \overline{1, T})$. На состояние системы может быть наложено ограничение, связанное с надежностью снабжения: $x_t \geqslant \overset{\circ}{x}$, где $\overset{\circ}{x}$ — величина некоторого страхового запаса, гарантирующего с заданной надежностью от сбоев в системе. Объединение ограничений, налагаемых на состояние системы и вектор управления, обозначим Ω. Получим задачу

$$\min Z = \sum_{t=1}^{T}(k\delta_t + S_t x_t), \quad \mathbf{u} \in \Omega.$$

Особенности задач динамического программирования. На основании приведенных примеров можно выделить типичные особенности многошаговых задач.

1. Рассматривается система, состояние которой на каждом шаге определяется вектором $\mathbf{x}_t$. Дальнейшее изменение ее состояния зависит только от данного состояния $\mathbf{x}_t$ и не зависит от того, каким путем система пришла в него. Такие процессы называются *процессами без последействия*.

2. На каждом шаге выбирается одно решение u_t, под действием которого система переходит из предыдущего состояния $\mathbf{x}_{t-1}$ в новое $\mathbf{x}_t$. Это новое состояние является функцией состояния на начало интервала $\mathbf{x}_{t-1}$ и принятого в начале интервала решения u_t, т. е.

$$\mathbf{x}_t = \mathbf{x}_t(\mathbf{x}_{t-1}, u_t).$$

3. Действие на каждом шаге связано с определенным выигрышем (доходом, прибылью) или потерей (издержками), которые зависят от состояния на начало шага (этапа) и принятого решения.

4. На векторы состояния и управления могут быть наложены ограничения, объединение которых составляет область допустимых решений $\mathbf{u} \in \Omega$.

5. Требуется найти такое допустимое управление u_t для каждого шага t, чтобы получить экстремальное значение функции цели за все T шагов.

Любую допустимую последовательность действий для каждого шага, переводящую систему из начального состояния в конечное, называют *стратегией управления*. Допустимая стратегия управления, доставляющая функции цели экстремальное значение, называется *оптимальной*.

Геометрическая интерпретация задачи динамического программирования состоит в следующем. Пусть n — размерность пространства состояний. В каждый момент времени координаты системы имеют вполне определенные значения. С изменением времени t могут изменяться значения координат вектора состояния. Назовем переход системы из одного состояния в другое траекторией ее движения в пространстве состояний. Такой переход осуществляется воздействием на координаты состояния. Пространство, в котором координатами служат состояния системы, называется фазовым. Особенно наглядно задачу динамического программирования можно интерпретировать в случае, если пространство состояний двухмерно.

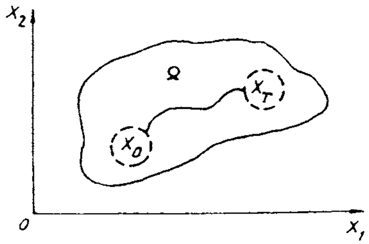


Рис. 8.1

Область возможных состояний в этом случае изобразится некоторой фигурой Ω, начальное и конечное состояния системы — точками $x_0, x_T \in \Omega$ (рис. 8.1). *Управление* — это воздействие, переводящее систему из начального состояния в конечное. Для многих экономических задач не известно начальное либо конечное состояние, а известна область X_0 или X_T, которой эти точки принадлежат. Тогда допустимые управления переводят точки из области X_0 в X_T. Задача динамического программирования геометрически может быть сформулирована следующим образом: найти такую фазовую траекторию, начинающуюся в области X_0 и оканчивающуюся в области X_T, для которой функция цели достигает экстремального значения. Если в задаче динамического программирования известны начальное и конечное состояния, то говорят о задаче с закрепленными концами. Если известны начальные и конечные области, то говорят о задаче со свободными концами.

8.2. ПРИНЦИПЫ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ. ФУНКЦИОНАЛЬНЫЕ УРАВНЕНИЯ БЕЛЛМАНА

Принцип оптимальности и погружения. Любую многошаговую задачу можно решать по-разному. Во-первых, можно считать неизвестными величинами u_t и находить экстремум целевой функции одним из существующих методов оптимизации, т. е. искать сразу все элементы решения на всех N шагах. Отметим, что этот путь не всегда приводит к цели, особенно когда целевая функция задана в виде таблиц или число переменных очень велико. Второй путь основан на идее проведения оптимизации поэтапно. Поэтапность отнюдь не предполагает изолированности в оптимизации этапов. Наоборот, управление на каждом шаге выбирается с учетом всех его последствий. Обычно второй способ оптимизации оказывается проще, чем первый, особенно при большом числе шагов. Идея постепенной, пошаговой оптимизации составляет суть метода динамического программирования. Оптимизация одного шага, как правило, проще оптимизации всего процесса в целом. Лучше много раз решать сравнительно простую задачу, чем один раз — сложную.

С первого взгляда идея может показаться тривиальной: если трудно оптимизировать сложную задачу, то следует разбить ее на ряд более простых. На каждом шаге оптимизируется задача меньшего размера, что уже нетрудно. При этом принцип динамического программирования вовсе не предполагает, что каждый шаг оптимизируется изолированно, независимо от других. Напротив, пошаговое управление должно выбираться с учетом всех его последствий.

Пусть, например, планируется работа группы промышленных предприятий, из которых одни заняты выпуском предметов потребления, а другие производят для этого машины. Задачей является получение за T лет максимального объема выпуска предметов потребления. Пусть планируются капиталовложения на первый год. Исходя из узких интересов этого года, мы должны были бы все средства вложить в производство предметов потребления, пустить имеющиеся машины на полную мощность и добиться к концу года максимального объема выпуска продукции. Однако относительно всего периода планирования такое решение будет нерациональным. Необходимо выделить часть средств на производство машин. При этом объем продукции за первый год снизится, зато будут созданы условия, позволяющие увеличить ее выпуск в последующие годы.

Приведем второй пример. Пусть прокладывается участок железнодорожного пути между пунктами A и B. Различные варианты трассы требуют неодинаковых затрат в связи с неоднородностью грунта, особенностями рельефа, естественными препятствиями и т. д. Требуется так провести дорогу из A в B, чтобы суммарные затраты на ее сооружение были минимальны.

Отметим, что в данной задаче нет естественного деления на шаги. Это деление вводится искусственно, для чего расстояние между A и B разбивается на N частей и за шаг оптимизации принимается каждая такая часть.

Таким образом, одним из условий применимости метода динамического программирования является возможность разбиения процесса оптимизации решения на ряд однотипных шагов (этапов), каждый из которых планируется отдельно, но с учетом состояния системы на начало этапа и последствий принятого решения. Однако из этого правила есть исключение. Среди всех шагов существует один, который может планироваться без учета последствий. Это последний шаг. Он может быть изучен и спланирован сам по себе наилучшим (в смысле выбранного критерия) образом, поскольку за ним нет больше этапов. Отсюда получаем одну из специфических особенностей динамического программирования: всю вычислительную процедуру программирования целесообразно разворачивать от конца к началу. Раньше всех планируется последний N-й шаг, за ним $(N-1)$-й и т. д. Но как найти оптимальное управление u_N на N-м шаге, если оно определяется не только целью управления, но и состоянием системы на начало этого шага? Сделать это можно на основе предположений об ожидаемых исходах предшествующего, но еще не исследованного этапа, т. е. о значениях x_{N-1}.

Для каждого возможного исхода x_{N-1} на $(N-1)$-м этапе находим оптимальное управление на N-м этапе. Такой набор оптимальных управлений, зависящих от возможных исходов предыдущего этапа, называется *условно-оптимальным решением* $u_N^*(x_{N-1})$. Завершив анализ конечного этапа, рассматривают аналогичную задачу для предпоследнего этапа, требуя, чтобы функция цели достигала экстремального значения на двух последних этапах вместе. Это дает условно-оптимальное решение на предпоследнем этапе $u_{N-1}^*(x_{N-2})$, т. е. делаются всевозможные предположения о том, чем кончился предыдущий $(N-2)$-й шаг, и для каждого из предположений находится такое управление на $(N-1)$-м шаге, при котором эффект за последние два шага (из них последний уже оптимизирован) будет максимален. Тем самым мы найдем для каждого исхода $(N-2)$-го шага условно-оптимальное управление на $(N-1)$-м и условно-оптимальное значение функции цели на последних двух шагах. Проделав такой поиск условно-оптимальных управлений для каждого шага от конца к началу, найдем последовательность условно-оптимальных управлений $u_1^*(x_0), u_2^*(x_1), \dots, u_N^*(x_{N-1})$.

Условно-оптимальные управления дают возможность найти не условное, а просто оптимальное управление на каждом шаге. В самом деле, пусть начальное состояние x_0 известно. Тогда, проделав процедуру движения от конца к началу, находим $u_1^*(x_0)$. Так как начальное состояние x_0 определяется однозначно, это оптимальное управление для первого шага. Вместе с тем находим экстремальное значение целевой функции относительно всего процесса. Зная оптимальное действие (с точки зрения всего процесса) для первого шага, выявим, к какому состоянию перейдет система в результате этого действия, т. е. найдем оптимальное состояние системы x_1^* на начало второго этапа. Но для всех возможных состояний на начало второго этапа выявлены оптимальные управления. Таким образом, зная x_1^*, установим оптимальное управление для второго этапа $u_2^*(x_1^*)$ и т. д. Проделав обратное движение по условно-оптимальным управлениям от начала к концу, найдем оптимальные управления для всех этапов.

Таким образом, в процессе оптимизации управления методом динамического программирования многошаговый процесс проходится дважды. Первый раз от конца к началу, в результате чего находятся условно-оптимальные управления и условно-оптимальное значение функции цели для каждого шага, в том числе оптимальное управление для первого шага и оптимальное значение функции цели для всего процесса. Второй раз — от начала к концу, в результате чего находятся уже оптимальные управления на каждом шаге с точки зрения всего процесса. Первый этап сложнее и длительнее второго, на втором остается лишь отобрать рекомендации, полученные на первом. Следует отметить, что понятия "конец" и "начало" можно поменять местами и разворачивать процесс оптимизации в другом направлении. С какого конца начать — диктуется удобством выбора этапов и возможных состояний на их начало.

Из качественного анализа идеи поэтапной оптимизации можно сформулировать следующие принципы, лежащие в основе динамического программирования: принцип оптимальности и принцип погружения.

Принцип оптимальности. Оптимальное управление на каждом шаге определяется состоянием системы на начало этого шага и целью управления. Или в развернутой форме: оптимальная стратегия обладает таким свойством, что, каковы бы ни были начальное состояние и начальные решения, последующие решения должны приниматься исходя из оптимальной стратегии с учетом состояния, вытекающего из первого решения. Ниже будет показано, что этот принцип имеет довольно простую математическую интерпретацию, выражающуюся в составлении определенных рекуррентных соотношений (функциональных уравнений Р. Беллмана).

Принцип погружения. Природа задачи, допускающей использование метода динамического программирования, не меняется при изменении количества шагов N, т. е. форма такой задачи инвариантна относительно N. В этом смысле всякий конкретный процесс с заданным числом шагов оказывается как бы погруженным в семейство подобных ему процессов и может рассматриваться с позиции более широкого класса задач.

Реализация названных принципов дает гарантию того, что решение, принимаемое на очередном шаге, окажется наилучшим относительно всего процесса в целом, а не узких интересов данного этапа. Последовательность пошаговых решений приводит к решению исходной N-шаговой задачи.

Функциональные уравнения Беллмана. Как отмечалось выше, в основе динамического программирования лежит принцип оптимальности, указывающий на процедуру построения оптимального управления. Так как оптимальной стратегией может быть только та, которая одновременно оптимальна и для любого количества оставшихся шагов, ее можно строить по частям: сначала для последнего этапа, затем для двух последних, для трех и т. д., пока не придем к первому шагу. Отсюда принцип оптимальности связан со вторым принципом — погружения, согласно которому при решении исходной задачи ее как бы погружают в семейство подобных ей и решают для одного последнего этапа, для двух последних и т. д., пока не получат решение исходной задачи.

Дадим математическую формулировку принципа оптимальности для задач с аддитивным критерием оптимальности (сепарабельная функция цели). Для простоты будем считать, что начальное x_0 и конечное x_T состояния системы заданы. Обозначим через $z_1(x_0, u_1)$ значение функции цели на первом этапе при начальном состоянии системы x_0 и при управлении u_1, через $z_2(x_1, u_2)$ — соответствующее значение функции цели только на втором этапе, ..., через $z_i(x_{i-1}, u_i)$ — на i-м этапе, ..., через $z_N(x_{N-1}, u_N)$ — на N-м этапе. Очевидно, что

$$Z = z(x_0, u) = \sum_{i=1}^{N} z_i(x_{i-1}, u_i). \tag{8.1}$$

Надо найти оптимальное управление $\mathbf{u}^* = (u_1^*; u_2^*; \dots; u_N^*)$, такое, что доставляет экстремум целевой функции (8.1) при ограничениях $\mathbf{u} \in \Omega$.

Для решения этой задачи погружаем ее в семейство подобных. Введем обозначения. Пусть Ω_N, $\Omega_{N-1,N}, \dots$, $\Omega_{\overline{1,N}} \equiv \Omega$ — соответственно области определения для подобных задач на последнем этапе, двух последних и т. д.; Ω — область определения исходной задачи. Обозначим через $F_1(x_{N-1})$, $F_2(x_{N-2})$, $\dots$, $F_k(x_{N-k})$, $\dots$, $F_N(x_0)$ соответственно условно-оптимальные значения функции цели на последнем этапе, двух последних и т. д., на k последних и т. д., на всех N этапах.

Начинаем с последнего этапа. Пусть x_{N-1} — возможные состояния системы на начало N-го этапа. Находим:

$$F_1(x_{N-1}) = \max_{u_N \in \Omega_N}(\min) z_N(x_{N-1}, u_N). \tag{8.2}$$

Для двух последних этапов получаем

$$F_2(x_{N-2}) = \max_{u_{N-1} \in \Omega_{N-1,N}}(\min)\ (z_{N-1}(x_{N-2}, u_{N-1}) + F_1(x_{N-1}))\,. \tag{8.3}$$

Аналогично:

$$F_3(x_{N-3}) = \max_{u_{N-2} \in \Omega_{N-2,N-1,N}}(\min)\ (z_{N-2}(x_{N-3}, u_{N-2}) + F_2(x_{N-2}))\,, \tag{8.4}$$

. . . .

$$\begin{aligned} F_k(x_{N-k}) = \max_{u_{N-k+1} \in \Omega_{N-k+1,\dots,N}}(\min)\ &(z_{N-k+1}(x_{N-k}, u_{N-k+1}) + \\ &+ F_{k-1}(x_{N-k+1})), \end{aligned} \tag{8.5}$$

. . . .

$$F_N(x_0) = \max_{u_1 \in \Omega}(\min)\,(z_1(x_0, u_1) + F_{N-1}(x_1))\,. \tag{8.6}$$

Выражение (8.6) представляет собой математическую запись принципа оптимальности. Выражение (8.5) — общая форма записи условно-оптимального значения функции цели для k оставшихся этапов. Выражения (8.2) — (8.6) называются *функциональными уравнениями Беллмана.* Отчетливо просматривается их рекуррентный (возвратный) характер, т. е. для нахождения оптимального управления на N шагах нужно знать условно-оптимальное управление на предшествующих $N-1$ этапах и т. д. Поэтому функциональные уравнения часто называют *рекуррентными (возвратными) соотношениями Беллмана.*

8.3. РЕШЕНИЕ ЭКОНОМИЧЕСКИХ ЗАДАЧ МЕТОДОМ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Пример 8.1 (задача о минимизации расхода горючего). Пусть самолет летит со скоростью v_0 на высоте h_0. Он должен подняться на высоту h_k, изменив скорость до v_k. Известен расход горючего при подъеме самолета с любой высоты h_1 на любую высоту $h_2 > h_1$ при постоянной скорости, а также расход горючего при увеличении скорости от любого значения v_1 до любого значения $v_2 > v_1$. Требуется найти оптимальное управление набором высоты и скорости, при котором общий расход горючего минимален.

Решение. Состояние системы описывается двумя координатами v и h. Начальное состояние $X_0(v_0; h_0)$ и конечное $X_к(v_к; h_к)$ системы известны. Имеем задачу с закрепленными концами. Выберем шаг дискретизации по v и h. Разобьем отрезки $v_к - v_0$ и $h_к - h_0$ на целое число шагов и будем считать, что на каждом этапе самолет может увеличивать либо высоту, либо скорость на один шаг. Пусть для определенности диапазоны изменения v и h разбиты соответственно на 5 и 4 интервала. Расход горючего на каждом шаге изменения v и h показан на рис. 8.2. Числа у вертикальных линий показывают его расход при наборе высоты с постоянной скоростью на данном этапе, а числа у горизонтальных линий — при увеличении скорости без изменения высоты.

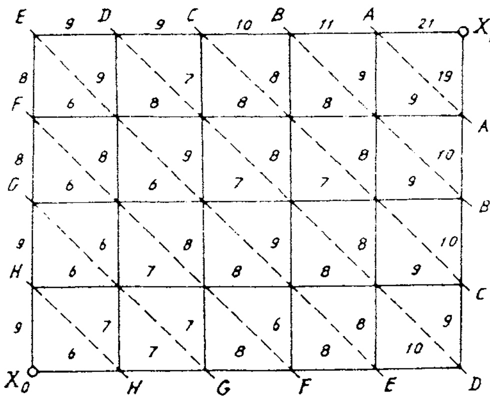


Рис. 8.2

Для решения задачи методом динамического программирования представим процесс как многоэтапный. На рис. 8.2 этапы выделены диагоналями $A - A$, $B - B$, $C - C$, $D - D$, $E - E$, $F - F$, $G - G$, $H - H$. В состояние $X_к$ можно попасть на последнем этапе ($A - A$) из точек A_1 и A_2 — возможных состояний на этом этапе (рис. 8.3). У точек $A_1(21)$ и $A_2(19)$ указан расход горючего для преодоления последнего этапа — перехода в конечное состояние $X_к(0)$. Нуль у точки $X_к$ означает, что после достижения точки $X_к$ горючее не расходуется. В точки A_1 и A_2 последнего этапа можно попасть из возможных состояний B_1, B_2, B_3 предпоследнего этапа $B - B$. В узлах $B_1(32)$, $B_2(28)$, $B_3(29)$ показан минимальный расход горючего для преодоления двух последних этапов. Например, из точки B_2 в $X_к$ можно попасть, двигаясь по вертикали либо по горизонтали к состояниям A_1 и A_2. Поэтому у точки B_2 поставлено число $28 = \min(9 + 21,\ 9 + 19)$ и т. д. На рис. 8.3 у каждой из узловых точек записано минимальное количество горючего для преодоления оставшихся этапов в случае, если самолет попадает в данное состояние этапа. У точки X_0 стоит число 75, показывающее минимальный расход горючего на всех этапах.

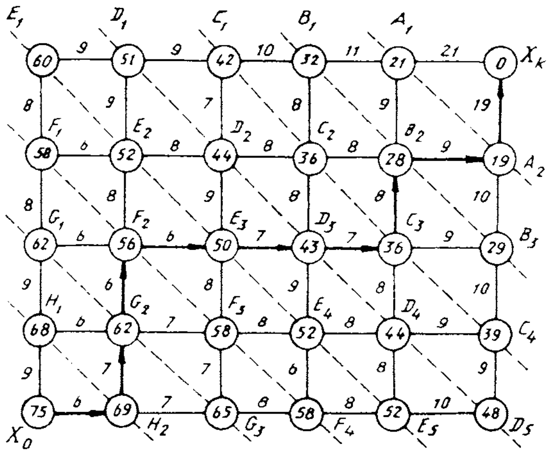


Рис. 8.3

На этом примере наглядно просматривается структура динамического программирования. Процесс разбивается на ряд шагов. Оптимизация начинается с последнего шага. На каждом шаге "попятного" движения находится условно-оптимальное управление. Условность его состоит в том, что на каждом этапе определяют оптимальное продолжение исходя из предположения, что самолет попал в данный узел (состояние). Управление на каждом шаге состоит в том, чтобы определить, как перемещаться на этом шаге — по горизонтали или по вертикали. Когда процесс "попятного" движения доведен до начального состояния X_0, находится уже не условно-оптимальное, а просто оптимальное управление, так как начальное состояние известно. Найдя минимальное значение целевой функции $\min Z = 75$ и оптимальное управление на первом этапе $(75 = 6 + 69)$ — $(X_0 - H_2)$, совершаем возвратное движение по всем этапам до конечного состояния $(X_0 - H_2 - G_2 - F_2 - E_3 - D_3 - C_3 - B_2 - A_2 - X_к)$. В этом случае на каждом шаге из множества условно-оптимальных управлений выбирается просто оптимальное. Если, например, приписать факту изменения скорости значение 0, а изменению высоты 1, то координаты вектора управления примут значение 0 или 1. Оптимальное управление запишется так:

$$\mathbf{u}^* = \begin{pmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 & u_9 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad z(\mathbf{u}^*) = 75.$$

В этом примере не учитывается возможность одновременного набора высоты и увеличения скорости. Одновременному увеличению высоты и скорости на рис. 8.4 соответствует движение по диагонали, идущей в направлении $X_{\text{к}}$. Проведем такие диагонали в каждом прямоугольнике и запишем над ними расход горючего. Теперь каждая узловая точка области возможных состояний, за исключением точек, лежащих на прямых $v_{\text{к}}$ и $h_{\text{к}}$, связана в направлении $X_{\text{к}}$ с тремя другими. Нумерацию этапов можно вести с конца к началу по линиям $A-A$, $B-B, \ldots$, а также по вертикальным прямым $X_{\text{к}} - D_5$, $A_1 - E_5, \ldots$ или по горизонтальным прямым $X_{\text{к}} - E_1$, $A_2 - F_1, \ldots$ Условно-оптимальные значения функции цели, записанные в узловых точках, и конечное оптимальное решение не будут зависеть от способа нумерации этапов.

На рис. 8.4 стрелками выделено оптимальное решение: $z(\mathbf{x}^*) = 72$.

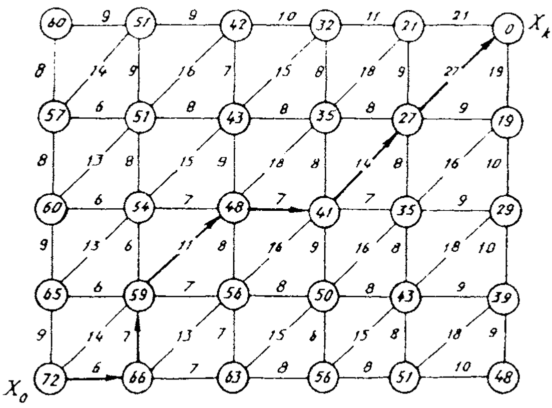


Рис. 8.4

Пример 8.2 (задача выбора кратчайшего пути). Задана транспортная сеть, на которой указаны пункт отправления A и пункт назначения B (рис. 8.5). Между ними имеется много других пунктов. Некоторые

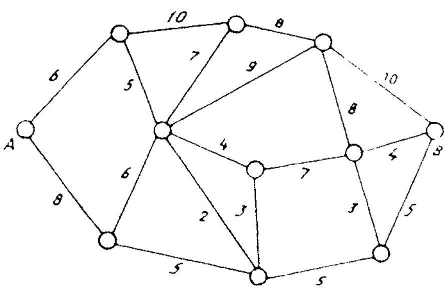


Рис. 8.5

соединены между собой участками пути. Над каждым участком транспортной сети поставлены цифры, указывающие расстояния между данными пунктами. Требуется составить маршрут минимальной длины из пункта A в пункт B.

Эта задача находит широкое применение в различных сферах деятельности. По сути решения аналогичными являются задачи выбора оптимальных схем транспортных перевозок, радиорелейных и телевизионных сетей, мелиорационных и ирригационных систем и т. п.

Выполнив разбиение сети на этапы и определив узлы (состояния на начало этапа), нужно найти в этих узлах минимальное расстояние от данного узла до точки B, если процесс динамического программирования разворачивается от конца к началу (справа налево), или минимальное расстояние от точки A до данной точки, если процесс разворачивается от начала к концу (слева направо).

Разбиение всей сети на этапы от A к B можно произвести двумя способами. Во-первых, можно произвольно провести вертикальные линии *1 – 1*, *2 – 2*, *3 – 3*, *4 – 4*,... и оценить отрезки путей, на которые они их делят (рис. 8.6). Во-вторых, можно упорядочить вершины по группам (см. пример 4.1) и, определив направление движения по каждому участку, развивать процесс динамического программирования от конца к началу или наоборот.

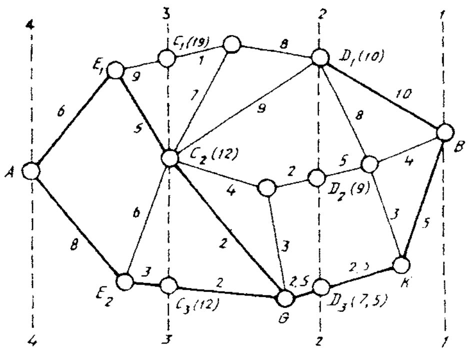


Рис. 8.6

При первом способе разбиения транспортной сети на этапы начнем процесс выбора кратчайшего пути от конца к началу (рис. 8.6). Найдем кратчайшие пути, соединяющие конечный пункт B с каждой точкой пересечения вертикали *2 – 2* с транспортной сетью. Таких точек пересечения три: D_1, D_2 и D_3. Для точки D_1 $\min(10; 8 + 4; 8 + 3 + 5) = 10$, для D_2 $\min(5 + 4; 5 + 3 + 5) = 9$, для D_3 $\min(2{,}5 + 3 + 4; 2{,}5 + 5) = 7,5$. На рис. 8.6 условно-оптимальные (кратчайшие расстояния от точек D_1, D_2 и D_3 до конечного пункта B) значения целевой функции показаны в скобках у состояний: $D_1(10)$, $D_2(9)$, $D_3(7{,}5)$. Далее рассматриваются точки пересечения вертикали $3-3$ с участками сети. Находим кратчайшие расстояния от этих точек до пункта B. Для C_1 $\min(1+8+10)=19$, для C_2 $\min(7+8+10; 9+10; 4+2+9; 2+2,5+7,5)=12$, для C_3 $\min(2+3+2+9; 2+2,5+7,5)=12$. Наконец, для точки A $\min(6+9+$ $+19;\ 6+5+12;\ 8+6+12;\ 8+3+12)=23$.

Таким образом, кратчайшее расстояние от A до B равно 23. Остается выбрать на сети кратчайший путь из A в B. Оптимальных участков этого пути на первом этапе два: AE_1C_2 и AE_2C_3. От этапа $3-3$ находим пути C_2GD_3 и C_3GD_3. От $2-2$ к $1-1$ ведет путь D_3KB. Итак, путей минимальной длины два, на рис. 8.6 они выделены жирной линией: $AE_1C_2GD_3KB$ и $AE_2C_3GD_3KB$.

Решим ту же задачу, разворачивая процесс динамического программирования от начала к концу. Вершины графа, приведенного на рис. 8.5, размещены по группам в предположении, что движение разрешено слева направо (рис. 8.7).

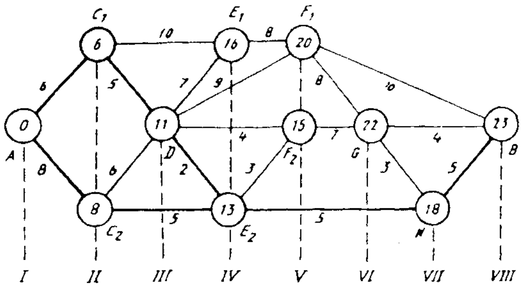


Рис. 8.7

Точки, стоящие на вертикалях $I-VIII$, — это состояния системы на начало следующего этапа. В точки C_1 и C_2 можно попасть из A, пройдя соответственно путь в 6 и 8 единиц. Эти расстояния показаны в кружках около точек C_1 и C_2. Переходим на вертикаль III. Состояния этой группы описываются одной точкой D. Теперь обозначим через $L(i)$ минимальный путь, предшествующий состоянию i, т. е. минимальный путь от A до i (n — число этапов). Тогда у точки D должно стоять $L_2(D)=\min(5+6;\ 6+8)=11$. Для состояний вертикали IV имеем: точка E_1 — $\min(10+6;\ 7+11)=16$; точка E_2 — $\min(2+11;\ 5+8)=13$. Вертикаль V: точка F_1 — $\min(8+16;\ 9+11)=20$; F_2 — $\min(4+11;\ 3+13)=15$. Для состояния вертикали VI: точка G — $\min(8+20;\ 7+15)=22$. Для состояния вертикали VII имеем: точка H — $\min(3+22;\ 5+13)=18$. Наконец, для B имеем $\min(10+20;\ 4+22;\ 5+18)=23$. Проходя теперь процесс в обратном порядке, выявляем значение функции цели и путь минимальной длины.

Впрочем, процесс можно развивать и от точки B к A. Очевидно, что второй подход имеет значительные преимущества перед первым: при первом подходе нужно производить дополнительную оценку отсекаемых вертикалями участков путей, что связано с анализом (неформальным) этих участков.

Пример 8.3 (задача распределения ресурсов). Пусть на реконструкцию и модернизацию основного производства объединению выделяется некоторый объем материальных ресурсов X_0. Имеется N предприятий, между которыми нужно распределить данный ресурс. Обозначим через $z_i(x_i)$ прибыль, которую приносит народному хозяйству выделение i-му предприятию x_i $(i = \overline{1, N})$ единиц ресурса. Предполагается, что размер прибыли зависит как от выделенного количества ресурса, так и от предприятия. Причем констатируется, что: прибыль, получаемая каждым предприятием, измеряется в одних и тех же единицах; прибыль, получаемая любым из предприятий, не зависит от того, какое количество этого ресурса выделено другим предприятиям; общая прибыль объединения состоит из прибылей отдельных предприятий. Исследования показывают, что функция $z_i(x_i)$, как правило, имеет вид, приведенный на рис. 8.8. Эта кривая обладает следующими особенностями: 1) небольшое количество выделенного ресурса не приносит сколько-нибудь ощутимого эффекта (прибыли); 2) для каждого предприятия существует момент, начиная с которого дальнейшее увеличение данного ресурса не эффективно.

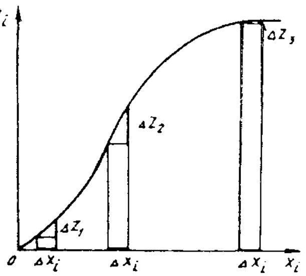


Рис. 8.8

Решение. Сформулированные выше предположения приводят к функции цели:

$$\max Z = z_1(x_1) + \ldots + z_N(x_N). \tag{8.7}$$

Задача оптимального распределения возникает оттого, что имеется ограниченный объем ресурса X_0, т. е.

$$x_1 + \ldots + x_N = X_0, \quad x_i \geqslant 0 \quad (i = \overline{1, N}). \tag{8.8}$$

Для решения задачи (8.7), (8.8) может быть применен аппарат классического математического анализа. Однако при более сложных ситуациях могут встретиться значительные трудности. Во-первых, обращение в нуль производной есть только необходимое, но не достаточное условие экстремума. Производная обращается в нуль не только во внутренних экстремальных точках, но и в точках перегиба. Во-вторых, в случае системы ограничений-неравенств, как правило, экстремум достигается на границе области допустимых решений, производная же обращается в нуль во внутренней точке. В-третьих, если система ограничений — дискретное множество или целевая функция разрывная либо вообще в некоторых точках недифференцируемая, то оптимизация таких задач требует новых методов, например динамического программирования.

Чтобы решить конкретную задачу распределения ограниченных ресурсов, применим аппарат функциональных уравнений Р. Беллмана. Погружаем ее в семейство подобных задач распределения. Вместо решения одной задачи с заданным объемом ресурса X_0 и фиксированным числом предприятий N рассмотрим их семейства, в которых объем выделенного ресурса x может меняться от нуля до X_0 и число предприятий — от 1 до N. Статическая задача распределения при таком подходе приобретает динамический характер. Введем последовательность функций $F_1(x)$, $F_2(x), \ldots, F_N(x)$, где $F_1(x)$ — это максимальная прибыль, если бы ресурс $0 \leqslant x \leqslant X_0$ был выделен только одному первому предприятию, $F_2(x)$ — максимальная прибыль, полученная при условии, что ресурс $0 \leqslant x \leqslant X_0$ был распределен двум предприятиям, и т. д. Пусть, наконец, $F_N(x)$ — максимальная прибыль, получаемая от распределения ресурса $0 \leqslant x \leqslant X_0$ между N предприятиями. Очевидно, что $F_N(x) = \max Z$.

В двух случаях элементы последовательности $F(x)$ имеют особенно простой вид: $F_i(0) = 0 \ \ (i = \overline{1, N})$, $\ F_1(x) = z_1(x)$, $\ 0 \leqslant x \leqslant X_0$.

Пусть ресурс $0 \leqslant x \leqslant X_0$ распределяется между двумя предприятиями. Если x_2 — объем ресурса, выделенного второму предприятию, то его прибыль составит $z_2(x_2)$. Оставшийся ресурс $x - x_2$ распределяется наилучшим образом. Общая прибыль для двух предприятий составит

$$F_2(x) = \max_{0 \leqslant x_2 \leqslant X_0} (z_2(x_2) + F_1(x - x_2)).$$

Рассуждая аналогично, найдем рекуррентное соотношение, связывающее $F_k(x)$ и $F_{k-1}(x - x_k)$ $\ (k = \overline{1, N})$ для произвольных значений $0 \leqslant x \leqslant X_0$. В самом деле, пусть $0 \leqslant x_k \leqslant X_0$ — количество ресурса, выделяемое для k-го предприятия. Тогда, каково бы ни было значение x_k, согласно принципу оптимальности, оставшийся ресурс $x - x_k$ распределится между остальными $k - 1$ предприятиями наилучшим образом. Так как $F_{k-1}(x - x_k)$ известно, то

$$F_k(x) = \max_{0 \leqslant x_k \leqslant X_0} (z_k(x_k) + F_{k-1}(x - x_k)).$$

Решение исходной задачи получим при $x = X_0$, $\ k = N$, т. е. из рекуррентного соотношения

$$F_N(X_0) = \max_{0 \leqslant x_N \leqslant X_0} (z_N(x_N) + F_{N-1}(X_0 - x_N)).$$

Найдя $F_N(X_0) = \max\limits_{x \in \Omega} z(X_0)$, определим x_N^*. Зная x_N^*, находим $X_0 - x_N^*$. Следовательно, $F_{N-1}(X_0 - x_N^*)$. Из выражения

$$F_{N-1}(X_0 - x_N^*) = \max_{0 \leqslant x_{N-1} \leqslant X_0 - x_N^*} (z_{N-1}(x_{N-1}) + F_{N-2}(X_0 - x_N^* - x_{N-1}))$$

находим x_{N-1}^* и т. д., т. е. процесс разворачивается в обратном направлении, при котором находятся уже не условно-оптимальные, а оптимальные значения функции цели на каждом этапе и оптимальная величина выделенного ресурса для одного, двух и более предприятий.

Вычислительная схема решения задачи распределения ресурсов методом динамического программирования состоит в следующем. Так как при решении функциональных уравнений Р. Беллмана невозможно протабулировать все значения функции цели для каждого предприятия $z_i(x)$ $(i = \overline{1, N})$, $0 \leqslant x \leqslant X_0$, то промежуток $0 \leqslant x \leqslant X_0$ разбивают, например, на n интервалов с шагом Δ и считают, что функции $z_i(x)$ и $F_i(x)$ определены только в точках $x = 0, \Delta, 2\Delta, 3\Delta, \ldots$, причем $n\Delta = X_0$. Значения $F_N(x)$ для x, отличных от точек $k\Delta$, где $k = 0, 1, 2, \ldots, n$, получают интерполяцией. При $i = 1$ функция $F_1(x)$ определяется равенством $F_1(x) = z_1(x)$. Множество значений $F_1(k\Delta) = z_1(k\Delta)$ $(k = 0, 1, 2, \ldots, n)$ записывают в таблицу. Зная значения $F_1(k\Delta)$, переходят к вычислению значений функции $F_2(k\Delta)$:

$$F_2(x) = \max_{0 \leqslant x_2 \leqslant X_0} (z_2(k\Delta) + F_1(X_0 - k\Delta)).$$

В ходе вычислений устанавливаются не только значения $F_2(x)$, $x = k\Delta$, $0 \leqslant x \leqslant X_0$ $(k = 0, 1, 2, \ldots, n)$, но и такие значения x_2, при которых достигает максимума выражение $z_2(k\Delta) + F_1(X_0 - k\Delta)$.

Переходим к отысканию функции $F_3(x)$ и т. д. Пройдя весь процесс вычисления функций $F_i(x)$ $(i = \overline{1, N})$, получим

$$F_N(X_0) = \max_{0 \leqslant x_N \leqslant X_0} (z_N(x_N) + F_{N-1}(X_0 - x_N)),$$

т. е. $F_N(X_0) = \max Z$, $x \in \Omega$. На последнем этапе достигается максимальное значение функции цели $F_N(X_0)$ и оптимальный объем ресурса, выделяемого N-му предприятию, т. е. x_N^*. Затем процесс вычислений просматривается в обратном порядке. Зная x_N^*, находим $X_0 - x_N^*$ — объем ресурса, подлежащего распределению между оставшимися $N - 1$ предприятиями. Так как раньше найдено значение

$$F_{N-1}(x) = \max_{0 \leqslant x \leqslant X_0} (z_{N-1}(x_{N-1}) + F_{N-2}(X_0 - x_{N-1})),$$

отсюда находим $F_{N-1}(X_0 - x_N^*)$ и, следовательно, x_{N-1}^* и т. д. Продолжая процесс к началу, определяем x_1^*. Тем самым будет завершено определение оптимального плана распределения ограниченного ресурса. Практическое применение рассмотренной схемы проиллюстрируем числовым примером.

Имеются четыре предприятия, между которыми следует распределить 400 единиц ограниченного ресурса. Получаемая предприятиями прибыль в зависимости от выделенной суммы x представлена в табл. 8.1. Приняв $\Delta = 80$, найти оптимальный план распределения.

Таблица 8.1

Выделяемый объем ресурса x	$z_1(x)$	$z_2(x)$	$z_3(x)$	$z_4(x)$
0	0	0	0	0
80	30	28	35	27
160	57	62	67	73
240	120	122	130	125
320	150	146	144	152
400	180	175	180	178

Пусть $N = 1$, тогда $F_1(x) = z_1(x)$.

Значения функции $F_1(x)$ помещены в табл. 8.2. Предположим теперь, что ресурс $X_0 = 400$ единиц распределяется между двумя предприятиями. Тогда $F_2(x) = \max(z_2(x_2) + F_1(x - x_2)),\ 0 \leqslant x \leqslant 400$.

Таблица 8.2

x \ $F_i(x)$	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$
0	0	0	0	0
80	30	30	35	35
160	57	62	67	72
240	120	122	130	130
320	150	152	160	160
400	180	182	192	203

Произведем вычисления значений функции $F_2(x)$ и представим их в табл. 8.2:

$$F_2(80) = \max_{x_2}(z_2(0) + F_2(80),\ z_2(80) + F_1(0)) = \max_{x_2}(0 + 30,\ 28 + 0) = 30,$$

$$F_2(160) = \max_{x_2}(z_2(0) + F_1(160), z_2(80) + F_1(80), z_2(160) + F_1(0)) =$$
$$= \max_{x_2}(0 + 57,\ 28 + 30,\ 62 + 0) = 62,$$

$$F_2(240) = \max_{x_2}(z_2(0) + F_1(240), z_2(80) + F_1(160), z_2(160) + F_1(80),$$
$$z_2(240) + F_1(0)) = \max_{x_2}(0 + 120,\ 28 + 57,\ 62 + 30,\ 122 + 0) = 122.$$

Аналогично имеем:

$$F_2(320) = \max_{x_2}(0 + 150,\ 28 + 120,\ 62 + 57,\ 122 + 30,\ 146 + 0) = 152,$$

$$F_2(400) = \max_{x_2}(0 + 180, 28 + 150, 62 + 120, 122 + 57, 146 + 30, 175 + 0) = 182.$$

Пусть, далее, имеющаяся сумма распределяется между тремя предприятиями. Получим $F_3(x) = \max\limits_{x_3}(z_3(x_3) + F_2(x - x_3)),\ 0 \leqslant x_3 \leqslant 400$.

Значения функции $F_3(x)$ представлены в табл. 8.2:

$$F_3(80) = \max_{x_3}(z_3(0) + F_2(80),\ z_3(80) + F_2(0)) = \max_{x_3}(0 + 30,\ 35 + 0) = 35,$$

$$F_3(160) = \max_{x_3}(z_3(0) + F_2(160), z_3(80) + F_2(80), z_3(160) + F_2(0)) =$$
$$= \max_{x_3}(0 + 62,\ 35 + 30,\ 67 + 0) = 67,$$

$$F_3(240) = \max_{x_3}(z_3(0) + F_2(240), z_3(80) + F_2(160), z_3(160) + F_2(80),$$
$$z_3(240) + F_2(0)) = \max_{x_3}(0 + 122,\ 35 + 62,\ 67 + 30,\ 130 + 0) = 130,$$

$$F_3(320) = \max_{x_3}(0 + 152,\ 35 + 122,\ 67 + 62,\ 130 + 30,\ 144 + 0) = 160,$$

$$F_3(400) = \max_{x_3}(0+182, 35+152, 67+122, 130+62, 144+30, 180+0) = 192.$$

Подключаем в процесс распределения четвертое предприятие объединения. Тогда

$$F_4(x) = \max_{x_4}(z_4(x) + F_3(x - x_4)), \quad 0 \leqslant x \leqslant 400.$$

Значения этой функции занесены в табл. 8.2:

$$F_4(80) = \max_{x_4}(0+35,\ 27+0) = 35,$$

$$F_4(160) = \max_{x_4}(0+67,\ 27+35,\ 72+0) = 72,$$

$$F_4(240) = \max_{x_4}(0+130,\ 27+67,\ 73+35,\ 25+0) = 130,$$

$$F_4(320) = \max_{x_4}(0+160,\ 27+130,\ 73+67,\ 125+35) = 160,$$

$$F_4(400) = \max_{x_4}(0+192, 27+160, 73+130, 125+67, 152+35, 178+0) = 203.$$

Из сводной таблицы 8.2 находим оптимальный план распределения. Максимальное значение функции цели составляет 203 единицы, т. е. $\max Z = 203$. Из приведенных расчетов значений $F_4(400)$ находим, что четвертому предприятию должно быть выделено $x_4^* = 160$ единиц ресурса, остальным трем — $400 - 160 = 240$ единиц. Из табл. 8.2 находим, что оптимальное распределение ресурса в 240 единиц между тремя предприятиями доставляет объединению 130 единиц прибыли. При этом третьему предприятию следует выделить $x_3^* = 240$ единиц ресурса. Отсюда $x_2^* = 0$, $x_1^* = 0$. Оптимальный план распределения

$$X^* = \begin{bmatrix} x_1^* & x_2^* & x_3^* & x_4^* \\ 0 & 0 & 240 & 160 \end{bmatrix}, \quad z(X^*) = 203.$$

Проверка очевидна: $z(X^*) = z_1(0) + z_2(0) + z_3(240) + z_4(160) = 0+$ $+0 + 130 + 73 = 203$.

Пример 8.4 (задача о замене оборудования). Задача о замене оборудования (обновлении, восстановлении, перестройке) имеет важное значение. Рассмотрим ее в упрощенной постановке. Известно, что оборудование со временем изнашивается, стареет физически и морально. В процессе эксплуатации, как правило, падает его производительность и растут эксплуатационные расходы на текущий ремонт. Со временем возникает необходимость замены оборудования, так как его дальнейшая эксплуатация обходится дороже, чем ремонт. Отсюда задача о замене может быть сформулирована так. В процессе работы оборудование дает ежегодно прибыль, требует эксплуатационных затрат и имеет остаточную стоимость. Эти характеристики зависят от возраста оборудования. В любом году оборудование можно сохранить, продать по остаточной цене и приобрести новое. В случае сохранения оборудования возрастают эксплуатационные расходы и снижается производительность. При замене нужны значительные дополнительные капитальные вложения. Задача состоит в определении оптимальной стратегии замен в плановом периоде, с тем чтобы суммарная прибыль за этот период была максимальной.

Для количественной формулировки задачи введем следующие обозначения: $r(t)$ — стоимость продукции, производимой за год на единице оборудования возраста t лет; $u(t)$ — расходы, связанные с эксплуатацией этого оборудования; $s(t)$ — остаточная стоимость оборудования возраста t лет; p — покупная цена оборудования; T — продолжительность планового периода; $t = 0, 1, 2, \ldots, T$ — номер текущего года.

Решение. Чтобы решить задачу, применим принцип оптимальности Р. Беллмана. Рассмотрим интервалы (годы) планового периода в последовательности от конца к началу. Введем функцию условно-оптимальных значений функции цели $F_k(t)$. Эта функция показывает максимальную прибыль, получаемую от оборудования возраста t лет за последние k лет планового периода. Здесь возраст оборудования рассматривается в направлении естественного хода времени. Например, $t = 0$ соответствует использованию совершенно нового оборудования. Временны́е же шаги процесса нумеруются в обратном порядке. Например, при $k = 1$ рассматривается последний год планового периода, при $k = 2$ — последние два года и т. д., при $k = T$ — последние T лет, т. е. весь плановый период. Направления изменения t и k показаны на рис. 8.9.

Рис. 8.9

В этой задаче систему составляет оборудование. Ее состояние характеризуется возрастом. Вектор управления это решение в момент $t = 0, 1, 2, \ldots, T$ о сохранении или замене оборудования. Для нахождения оптимальной политики замен следует проанализировать, согласно принципу оптимальности, процесс от конца к началу. Для этого сделаем предположение о состоянии оборудования на начало последнего года ($k = 1$). Пусть оборудование имеет возраст t лет. В начале T-го года имеются две возможности: 1) сохранить оборудование на T-й год, тогда прибыль за последний год составит $r(t) - u(t)$; 2) продать оборудование по остаточной стоимости и купить новое, тогда прибыль за последний год будет равна $s(t) - p + r(0) - u(0)$, где $r(0)$ — стоимость продукции, выпущенной на новом оборудовании за первый год его ввода; $u(0)$ — эксплуатационные расходы в этом году. Здесь целесообразно разворачивать процесс от конца к началу (см. рис. 8.9). Для последнего года ($k = 1$) оптимальной политикой с точки зрения всего процесса будет политика, обеспечивающая максимальную прибыль только за последний год. Учитывая значение прибыли при различном образе действия (замена — сохранение), приходим к выводу, что решение о замене оборудования возраста

t лет следует принять в случае, когда прибыль от нового оборудования на последнем периоде больше, чем от старого, т. е. при условии

$$s(t) - p + r(0) - u(0) > r(t) - u(t).$$

Если же

$$s(t) - p + r(0) - u(0) \leqslant r(t) - u(t),$$

то старое оборудование целесообразно сохранить.

Итак, для последнего года оптимальная политика и максимальная прибыль $F_1(t)$ находятся из условия

$$F_1(t) = \max_t \begin{cases} r(t) - u(t) & \text{(сохранение)}, \\ s(t) - p + r(0) - u(0) & \text{(замена)}. \end{cases}$$

Пусть $k = 2$, т. е. рассмотрим прибыль за два последних года. Делаем предположение о возможном состоянии t оборудования на начало предпоследнего года. Если в начале этого года принять решение о сохранении оборудования, то к концу года будет получена прибыль $r(t) - u(t)$. На начало последнего года оборудование перейдет в состояние $t + 1$, и при оптимальной политике в последнем году оно принесет прибыль, равную $F_1(t + 1)$. Таким образом, общая прибыль за два года составит $r(t) - u(t) + F_1(t + 1)$. Если же в начале предпоследнего года будет принято решение о замене оборудования, то прибыль за предпоследний год составит $s(t) - p + r(0) - u(0)$. Поскольку приобретено новое оборудование, на начало последнего года оно будет в состоянии $t = 1$. Следовательно, общая прибыль за последние два года при оптимальной политике в последнем году составит

$$s(t) - p + r(0) - u(0) + F_1(1).$$

Условно-оптимальной в последние два года будет политика, доставляющая максимальную прибыль:

$$F_2(t) = \max_t \begin{cases} r(t) - u(t) + F_1(t + 1) & \text{(сохранение)}, \\ s(t) - p + r(0) - u(0) + F_1(1) & \text{(замена)}. \end{cases}$$

Аналогично находим выражения для условно-оптимальной прибыли за три последних года, четыре и т. д. Общее функциональное уравнение примет вид

$$F_k(t) = \max_t \begin{cases} r(t) - u(t) + F_{k-1}(t + 1) & \text{(сохранение)}, \\ s(t) - p + r(0) - u(0) + F_{k-1}(1) & \text{(замена)}. \end{cases}$$

При $k = T$ получим $\max Z = F_T(t_0)$, причем

$$F_T(t_0) = \max_t \begin{cases} r(t_0) - u(t_0) + F_{T-1}(t_0 + 1) & \text{(сохранение)}, \\ s(t_0) - p + r(0) - u(0) + F_{T-1}(1) & \text{(замена)}. \end{cases}$$

Таким образом, разворачивая весь процесс от конца к началу, получаем, что максимальная прибыль за плановый период T составит $F_T(t_0)$. Так как начальное состояние t_0 известно, из выражения для $F_T(t_0)$ находим оптимальное решение в начале первого года, потом вытекающее из него оптимальное решение для второго года и т. д. Обратимся к числовому примеру.

Разработать оптимальную политику замены оборудования при условиях: 1) стоимость $r(t)$ продукции, производимой с использованием оборудования за год, и расходы $u(t)$, связанные с эксплуатацией оборудования, заданы табл. 8.3; 2) ликвидационная стоимость машины не зависит от ее возраста и равна 2; 3) цена нового оборудования со временем не меняется и равна 15; 4) продолжительность планового периода 12 лет. Итак, $s(t) = 2,\ \ p = 15,\ \ T = 12$.

Таблица 8.3

t / $r,\ u$	0	1	2	3	4	5	6	7	8	9	10	11	12
$r(t)$	30	30	29	29	29	28	28	27	27	26	24	23	23
$u(t)$	10	10	12	13	14	15	16	17	18	19	20	22	23

Запишем функциональные уравнения для $F_1(t)$ и $F_k(t)$ при числовых значениях нашего примера:

$$F_1(t) = \max_t \begin{cases} r(t) - u(t), \\ s(t) - p + r(0) - u(0) \end{cases} = \max_t \begin{cases} r(t) - u(t), \\ 2 - 15 + 30 - 10 \end{cases} =$$

$$= \max_t \begin{cases} r(t) - u(t) \quad (\text{сохранение}), \\ 7 \quad (\text{замена}), \end{cases} \tag{8.9}$$

$$F_k(t) = \max_t \begin{cases} r(t) - u(t) + F_{k-1}(t+1) \quad (\text{сохранение}), \\ 7 + F_{k-1}(1) \quad (\text{замена}). \end{cases} \tag{8.10}$$

Пользуясь выражениями (8.9), (8.10), будем последовательно вычислять значения максимальной прибыли $F_k(t)$ и записывать их в специальную таблицу (табл. 8.4). Первую строку получим, придавая параметру t в равенстве (8.9) значения $0, 1, \ldots, 12$ и используя исходные данные табл. 8.3. Например, при $t = 0$

$$F_1(0) = \max \begin{cases} r(0) - u(0), \\ 7 \end{cases} = \max \begin{cases} 30 - 10, \\ 7 \end{cases} = 20 \quad (\text{сохранение}).$$

Аналогично расчет ведется до $t = 9$:

$$F_1(9) = \max \begin{cases} 26 - 19, \\ 7 \end{cases} = 7 \quad (\text{сохранение}).$$

Заметим, что если прибыль от нового оборудования равна прибыли от старого, то старое лучше сохранить еще на год:

$$F_1(10) = \max \begin{cases} 24 - 20, \\ 7 \end{cases} = 7 \quad (\text{замена}).$$

Таблица 8.4

F_k \ t	0	1	2	3	4	5	6	7	8	9	10	11	12
$F_1(t)$	20	20	17	16	15	13	12	10	9	7	7	7	7
$F_2(t)$	40	37	33	31	28	27	27	27	27	27	27	27	27
$F_3(t)$	57	53	48	44	44	44	44	44	44	44	44	44	44
$F_4(t)$	73	68	61	60	60	60	60	60	60	60	60	60	60
$F_5(t)$	88	81	77	76	75	75	75	75	75	75	75	75	75
$F_6(t)$	101	97	93	91	90	88	88	88	88	88	88	88	88
$F_7(t)$	117	113	108	106	104	104	104	104	104	104	104	104	104
$F_8(t)$	133	128	123	120	120	120	120	120	120	120	120	120	120
$F_9(t)$	148	143	137	136	135	135	135	135	135	135	135	135	135
$F_{10}(t)$	166	157	153	151	150	150	150	150	150	150	150	150	150
$F_{11}(t)$	177	173	168	166	165	164	164	164	164	164	164	164	164
$F_{12}(t)$	193	188	183	181	180	180	180	180	180	180	180	180	180

Из табл. 8.3 видно, что $r(t) - u(t)$ с ростом t убывает. Поэтому при $t > 9$ оптимальной будет политика замены оборудования. Чтобы различать, в результате какой политики получается условно-оптимальное значение прибыли, будем эти значения (до $t = 9$ включительно оптимальной является политика сохранения) разграничивать жирной линией. Для заполнения второй строки табл. 8.4 используем формулу (8.10). Для $k = 2$ получаем

$$F_2(t) = \max_t \begin{cases} r(t) - u(t) + F_1(t+1), \\ 7 + F_1(1) \end{cases} =$$

$$= \max_t \begin{cases} r(t) - u(t) + F_1(t+1) & \text{(сохранение)}, \\ 27 & \text{(замена)}. \end{cases}$$

Придадим параметру t значения $0, 1, 2, \ldots, 12$, значения $r(t)$ и $u(t)$ возьмем из табл. 8.3, а значения $F_1(t+1)$ — из первой строки табл. 8.4. Для третьей строки расчетную формулу получим из равенства (8.10) при $k = 3$:

$$F_3(t) = \max_t \begin{cases} r(t) - u(t) + F_2(t+1), \\ 7 + F_2(1) \end{cases} =$$

$$= \max_t \begin{cases} r(t) - u(t) + F_2(t+1) & \text{(сохранение)}, \\ 44 & \text{(замена)} \end{cases}$$

и т. д. Заполнив табл. 8.4, данные ее используем для решения поставленной задачи. Эта таблица содержит много ценной информации и позволяет решать все семейство задач, в которое мы погружали исходную задачу.

Пусть, например, в начале планового периода имеем оборудование возраста 6 лет. Разработаем "политику замен" на двенадцатилетний период, доставляющую максимальную прибыль. Информация для этого имеется в табл. 8.4. Максимальная прибыль, которую можно получить за 12 лет при условии, что вначале имелось оборудование возраста 6 лет, находится в табл. 8.4 на пересечении столбца $t = 6$ и строки $F_{12}(t)$; она составляет 180 единиц.

Значение максимальной прибыли $F_{12}(6) = 180$ записано справа от ломаной линии, т. е. в области "политики замены". Это значит, что для достижения в течение 12 лет максимальной прибыли в начале первого года оборудование надо заменить. В течение первого года новое оборудование постареет на год, т. е., заменив оборудование и проработав на нем 1 год, мы за 11 лет до конца планового периода будем иметь оборудование возраста 1 год. Из табл. 8.4 берем $F_{11}(1) = 173$. Это значение располагается в области "политики сохранения", т. е. во втором году планового периода надо сохранить оборудование возраста 1 год, и, проработав на нем год, за 10 лет до конца планового периода будем иметь оборудование возраста 2 года.

Выясняем, что значение $F_{10}(2) = 153$ помещено в области сохранения. Работаем на оборудовании еще год. Теперь до конца планового периода осталось 9 лет, а возраст оборудования составляет 3 года. Находим $F_9(3) = 136$. Это область сохранения. Работаем на оборудовании еще год. Его возраст становится равным 4 годам. До конца планового периода остается 8 лет. Определяем $F_8(4) = 120$. Это область замен. Заменяем оборудование на новое. Проработаем на нем в течение четвертого года. Оно постареет на год. До конца планового периода останется 7 лет. Находим $F_7(1) = 113$. Это область сохранения. Продолжив подобные рассуждения, установим, что $F_6(2) = 93$, $F_5(3) = 76$ расположены в области сохранения, $F_4(4) = 60$ — в области замен, $F_3(1) = 53$, $F_2(2) = 33$, $F_1(3) = 16$ — в области сохранения. Разработанную политику изобразим следующей цепочкой:

$$F_{12}(6) \xrightarrow[\text{замена}]{\text{1-й год}} F_{11}(1) \xrightarrow[\text{сохранение}]{\text{2-й год}} F_{10}(2) \xrightarrow[\text{сохранение}]{\text{3-й год}}$$

$$F_9(3) \xrightarrow[\text{сохранение}]{\text{4-й год}} F_8(4) \xrightarrow[\text{замена}]{\text{5-й год}} F_7(1) \xrightarrow[\text{сохранение}]{\text{6-й год}}$$

$$F_6(2) \xrightarrow[\text{сохранение}]{\text{7-й год}} F_5(3) \xrightarrow[\text{сохранение}]{\text{8-й год}} F_4(4) \xrightarrow[\text{замена}]{\text{9-й год}}$$

$$F_3(1) \xrightarrow[\text{сохранение}]{\text{10-й год}} F_2(2) \xrightarrow[\text{сохранение}]{\text{11-й год}} F_1(3) \xrightarrow[\text{сохранение}]{\text{12-й год}} .$$

Таким образом, вместо поиска оптимальной "политики замен" на плановый период в 12 лет мы погрузили исходную задачу в семейство подобных, когда период меняется от 1 до 12. Решение ведется по принципу оптимальности для любого состояния системы, независимо от ее предыстории. Оптимальная "политика замен" является оптимальной на оставшееся число лет.

Табл. 8.4 содержит информацию для решения и других задач. Из нее можно найти оптимальную стратегию замены оборудования с любым начальным состоянием от 0 до 12 лет и на любой плановый период, не превосходящий 12 лет. Например, найдем "политику замен" на плановый период в 10 лет, если вначале имелось оборудование пятилетнего возраста:

$$F_{10}(5) \xrightarrow[\text{замена}]{\text{1-й год}} F_9(1) \xrightarrow[\text{сохранение}]{\text{2-й год}} F_8(2) \xrightarrow[\text{сохранение}]{\text{3-й год}}$$

$$F_7(3) \xrightarrow[\text{сохранение}]{\text{4-й год}} F_6(4) \xrightarrow[\text{сохранение}]{\text{5-й год}} F_5(5) \xrightarrow[\text{замена}]{\text{6-й год}}$$

$$F_4(1) \xrightarrow[\text{сохранение}]{\text{7-й год}} F_3(2) \xrightarrow[\text{сохранение}]{\text{8-й год}} F_2(3) \xrightarrow[\text{сохранение}]{\text{9-й год}}$$

$$F_1(4) \xrightarrow[\text{сохранение}]{\text{10-й год}} .$$

Задачу о замене оборудования мы упростили. На практике же деталями не пренебрегают. Легко учесть, например, случай, когда остаточная стоимость оборудования $s(t)$ зависит от времени. Может быть принято решение о замене оборудования не новым, а уже проработавшим некоторое время. Не составляет также труда учесть возможность капитального ремонта старого оборудования. При этом в понятие "состояние" системы необходимо включить время последнего ремонта оборудования. Функция $F_k(t_1, t_2)$ выражает прибыль за последние k лет планового периода при условии, что вначале имелось оборудование возраста t_1, прошедшее капитальный ремонт после t_2 лет службы. Характеристики r, s и u также будут функциями двух переменных t_1 и t_2.

9. ПАРАМЕТРИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

9.1. ПРОИЗВОДСТВЕННЫЕ ПРОБЛЕМЫ, ПРИВОДЯЩИЕ К ЗАДАЧАМ ЛИНЕЙНОГО ПАРАМЕТРИЧЕСКОГО ПРОГРАММИРОВАНИЯ

В математических моделях оптимизационных задач, которые рассматривались в этой книге, все коэффициенты были постоянными числами. Однако в экономической практике нередко возникают задачи, при формализации которых необходимо учитывать зависимость коэффициентов модели от некоторых параметров. Меняться могут либо коэффициенты целевой функции, либо компоненты вектора ограничений, либо коэффициенты системы ограничений, либо одновременно все перечисленные элементы. В этом случае приходится изучать поведение оптимального плана при изменении исходных данных в зависимости от параметров.

Раздел математического программирования, в котором рассматриваются экстремальные задачи с целевыми функциями и ограничениями, зависящими от параметров, разрабатываются численные методы, позволяющие находить оптимальный план сразу для совокупности значений параметров, и изучается поведение оптимальных планов этих задач при изменении параметров, называется *параметрическим программированием.*

Исследуем некоторые производственные проблемы, приводящие к *задачам линейного параметрического программирования* (ЗЛПП). В планово-производственных и экономических ситуациях, сводящихся к задачам линейного программирования вида

$$\max(\min) f = \sum_{j=1}^{n} c_j x_j; \tag{9.1}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant a_{i0} \quad (i = \overline{1, m}), \tag{9.2}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}), \tag{9.3}$$

предполагалось, что все c_j, a_{ij} и a_{i0} являются постоянными.

Допустим теперь, что в задаче с практическим содержанием (например, в задаче планирования сельскохозяйственного производства) коэффициенты c_j целевой функции (9.1) выражают цену единицы продукции j-го вида, а x_j — количество единиц j-й продукции, планируемое к реализации. Если цена на производимую продукцию носит сезонный характер, то она будет функцией времени t. При этом для продукции каждого вида цена может изменяться по-разному, что характеризуется определенным коэффициентом α_j. В подобной ситуации часть готовой продукции придется хранить в течение какого-то времени. В связи с этим стоимость продукции, описываемая целевой функцией, будет складываться из постоянной части $\sum\limits_j c_j x_j$ — стоимости продукции на момент изготовления — и переменной части $\sum\limits_j \alpha_j t x_j$, зависящей от срока хранения t.

Предположим, что зависимость цены от параметра t линейна, а ограничения задачи имеют вид неравенств (9.2), (9.3), в которых все a_{ij} и a_{i0} постоянны. Тогда придем к ЗЛПП вида

$$\max(\min) f_t = \sum_{j=1}^{n} (c_j + \alpha_j t) x_j; \tag{9.4}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant a_{i0} \quad (i = \overline{1, m}), \tag{9.5}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}), \tag{9.6}$$

$$t \in I_1,$$

где $\overline{\alpha} = (\alpha_1; \ldots; \alpha_n)$ — заданный вектор; I_1 — заранее выбранный интервал.

К подобным задачам мы придем и при решении проблемы производственного планирования, если c_j означает количество некоторого продукта, производимого в единицу времени, когда производство ведется по j-й технологии, а x_j — отрезок времени работы по этой технологии при условии, что изменение интенсивности производства по каждой технологии линейно зависит от некоторого параметра t (например, от качества перерабатываемого сырья или времени). В этой ситуации функция (9.4) выражает общее количество произведенного продукта при плане $(x_1; \ldots; x_n)$ использования различных технологий в зависимости от параметра t с соблюдением ограничений (9.5), (9.6).

Предположим теперь, что в задаче производственного планирования вида (9.1) — (9.3) могут изменяться в зависимости от параметра t (времени, используемой технологии, вместимости складских помещений и т. п.) запасы a_{i0} ресурсов, а остальные характеристики задачи (c_j, a_{ij}) постоянны. Тогда получим ЗЛПП вида

$$\max(\min) f = \sum_{j=1}^{n} c_j x_j;$$

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant a_{i0} + \gamma_i t \quad (i = \overline{1, m}),$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}),$$

$$t \in I_2,$$

где $\overline{\gamma} = (\gamma_1; \ldots; \gamma_m)$ — заданный вектор; I_2 — заданный интервал.

Встречаются производственные ситуации, при математическом описании которых возникает необходимость рассматривать только элементы матрицы $[a_{ij}]_{m \times n}$ технологических коэффициентов в зависимости от параметра t, а также ситуации, приводящие к ЗЛПП, в которых от одного параметра линейно зависят все характеристики задачи: c_j, a_{ij}, a_{i0}. В последнем случае задача принимает вид

$$\max(\min) f_t = \sum_{j=1}^{n} (c_j + \alpha_j t) x_j;$$

$$\sum_{j=1}^{n} (a_{ij} + \beta_{ij} t) x_j \leqslant a_{i0} + \gamma_i t \quad (i = \overline{1, m}),$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}),$$

где допустимые области изменения параметра t, а также постоянные $\alpha_j, \beta_{ij}, \gamma_i$ устанавливаются по производственным соображениям.

9.2. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ С ЦЕЛЕВОЙ ФУНКЦИЕЙ, ЗАВИСЯЩЕЙ ОТ ПАРАМЕТРА

Реальные экономические и производственные процессы бывают настолько сложными, что при их математическом моделировании приходится учитывать влияние на характеристики задачи не одного, а нескольких параметров. Исследование ЗЛПП со многими параметрами может оказаться весьма трудным. Ограничимся рассмотрением наиболее простого и достаточно хорошо изученного случая, когда от одного параметра зависят коэффициенты только целевой функции. Такую задачу можно сформулировать следующим образом: даны линейная при каждом $t \in [\alpha; \beta]$ целевая функция

$$f_t = \sum_{j=1}^{n}(c_j + d_j t)x_j \tag{9.7}$$

и система ограничений

$$\sum_{j=1}^{n} a_{ij}x_j = a_{i0} \quad (i = \overline{1, m}), \tag{9.8}$$

$$x_j \geqslant 0 \quad (j = \overline{1, n}) \tag{9.9}$$

(c_j, d_j, a_{ij}, a_{i0} — заданные постоянные), определяющая некоторый непустой многогранник M. Требуется разбить отрезок $[\alpha; \beta]$ на конечное число отрезков так, чтобы для всех значений параметра t из каждого частичного отрезка максимальное значение f_t достигалось в одной и той же вершине многогранника M.

Решение задачи (9.7) — (9.9) сводится к следующему.

1. Полагают $t = \alpha$ и решают обычную задачу линейного программирования, т. е. задачу отыскания максимума целевой функции

$$f_\alpha = \sum_{j=1}^{n}(c_j + \alpha d_j)x_j$$

при ограничениях (9.8), (9.9), которые по предположению совместны. В результате находят оптимальную вершину.

2. Определяют все значения параметра t, для которых $\max f_t$ достигается в той же вершине. Найденный отрезок исключают из исходного отрезка $[\alpha;\beta]$. Для оставшегося отрезка снова решают ЗЛП, т. е. возвращаются к п. 1. Так продолжается до тех пор, пока весь отрезок $[\alpha;\beta]$ не будет разделен на частичные отрезки. Поскольку на каждом шаге осуществляется переход от одной вершины многогранника M к другой, а число их конечно и они не могут повторяться, описанная процедура конечна.

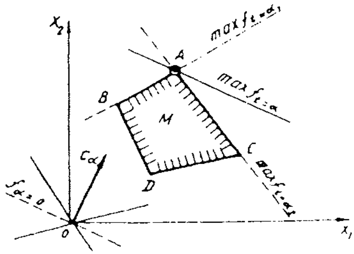


Рис. 9.1

С геометрический точки зрения система ограничений (9.8), (9.9) при $n = 2$ определяет некоторый многоугольник M, а уравнение $f_t = 0$ — пучок прямых, проходящих через начало координат. Положив $t = \alpha$, выделим одну из прямых $f_\alpha = 0$, занимающую на плоскости x_1Ox_2 вполне определенное положение (рис. 9.1). Перемещая эту прямую в направлении вектора $\mathbf{c}_\alpha$, находим вершину A многоугольника M, в которой функция f_t достигает максимума (при $t = \alpha$). Для получения всех значений t, при которых функция f_t достигает максимума в вершине A, надо поворачивать прямую $\max f_{t=\alpha}$ вокруг точки A до совмещения с граничными прямыми AB $(t = \alpha_1)$ и AC $(t = \alpha_2)$. Если t будет, например, больше α_2, то получим оптимальный план, но уже в вершине C, которая имеет свой интервал изменения t.

Данная геометрическая интерпретация позволяет решать задачи вида (9.7) — (9.9) для $n = 2$ графическим способом.

При аналитическом решении данной задачи в соответствии с изложенной схемой полагаем $t = \alpha$. Тогда выражение (9.7) принимает вид

$$f_\alpha = \sum_{j=1}^{n} (c_j + \alpha d_j) x_j$$

и мы приходим к обычной ЗЛП. Для ее решения составляем симплекс-таблицу, к которой добавляем снизу еще две строки (табл. 9.1). В первой из них записываем числа c_j, а во второй — числа d_j. Дополнительные строки позволяют получать

Таблица 9.1

	1	$-x_1$	...	$-x_j$	...	$-x_n$
$0 =$	a_{10}					
...	...			(a_{ij})		
$0 =$	a_{m0}					
$f_\alpha =$	0	$-(c_1 + \alpha d_1)$	...	$-(c_j + \alpha d_j)$	...	$-(c_n + \alpha d_n)$
f_t	0	$-c_1$	...	$-c_j$	...	$-c_n$
	0	$-d_1$	...	$-d_j$	...	$-d_n$

коэффициенты целевой функции f_t: коэффициенты последней строки умножают на t и складывают с коэффициентами предпоследней. Коэффициенты целевой функции f_t потребуются для определения пределов изменения параметра t.

Решая записанную в табл. 9.1 (здесь мы пользуемся жордановыми таблицами) ЗЛП симплекс-методом, находим оптимальный план $\mathbf{x}_1^*$, т. е. вершину, в которой достигается оптимум (табл. 9.2).

Поскольку в табл. 9.2 содержится оптимальный план, то

$$b_{0j} + \alpha b'_{0j} \geqslant 0 \quad (j = \overline{m+1, n}).$$

Теперь надо определить t, при которых $\max f_t$ достигается в той же вершине, т. е. для которой записанный в табл. 9.2

план будет оставаться оптимальным. Но план оптимален, когда все элементы f_t-строки неотрицательны, т. е.

$$b_{0j} + tb'_{0j} \geqslant 0 \quad (j = \overline{m+1, n}). \tag{9.10}$$

Таблица 9.2

БП	1	СП				
		$-x_{m+1}$	$\dots$	$-x_{m+s}$	$\dots$	$-x_n$
$x_1 =$	b_{10}					
$\dots$	$\dots$			(b_{ij})		
$x_m =$	b_{m0}					
$f_\alpha =$	$b_{00} + \alpha b'_{00}$	$b_{0,m+1} + \alpha b'_{0,m+1}$	$\dots$	$b_{0s} + \alpha b'_{0s}$	$\dots$	$b_{0n} + \alpha b'_{0n}$
f_t	b_{00}	$b_{0,m+1}$	$\dots$	b_{0s}	$\dots$	b_{0n}
	b'_{00}	$b'_{0,m+1}$	$\dots$	b'_{0s}	$\dots$	b'_{0n}

Решив систему (9.10)

$$\left.\begin{array}{lcr} b_{0,m+1} & + \; tb'_{0,m+1} & \geqslant 0, \\ \dots\dots\dots & \dots\dots\dots & \dots\dots \\ b_{0j} & + \; tb'_{0j} & \geqslant 0, \\ \dots\dots\dots & \dots\dots\dots & \dots\dots \\ b_{0n} & + \; tb'_{0n} & \geqslant 0, \end{array}\right\} \tag{9.11}$$

можно найти все интересующие нас значения параметра t. При этом могут встретиться следующие случаи.

1. Все $b'_{0j} > 0$. Тогда неравенства системы (9.11) можно записать в виде $t \geqslant -b_{0j}/b'_{0j}$ $(j = \overline{m+1, n})$. Система будет удовлетворяться при всех $t \geqslant \max(-b_{0j}/b'_{0j})$. Обозначим $\alpha_1 = \max(-b_{0j}/b'_{0j})$, $\alpha_2 = +\infty$.

Итак, если $b'_{0j} > 0$, то в полуинтервале $[\alpha_1; \alpha_2)$ целевая функция достигает максимума в той же вершине $\mathbf{x}_1^*$.

2. Все $b'_{0j} < 0$. Аналогично предыдущему получаем, что в полуинтервале $(\alpha_1; \alpha_2]$, где $\alpha_1 = -\infty$, $\alpha_2 = \min(-b_{0j}/b'_{0j})$, целевая функция f_t достигает максимума в той же вершине $\mathbf{x}_1^*$.

3. Среди чисел b'_{0j} имеются как положительные, так и отрицательные. Разделим неравенства (9.11) на две группы: с

положительными b'_{0j} и отрицательными b'_{0j}. Неравенства первой группы удовлетворяются при $t \geqslant \max(-b_{0j}/b'_{0j})$, второй — при $t \leqslant \min(-b_{0j}/b'_{0j})$. Вся система будет удовлетворяться при $\max(-b_{0j}/b'_{0j}) \leqslant t \leqslant \min(-b_{0j}/b'_{0j})$.

Итак, целевая функция f_t достигает максимума в вершине $\mathbf{x}_1^*$ при всех $t \in [\alpha_1; \alpha_2]$, где $\alpha_1 = \max(-b_{0j}/b'_{0j})$, $\alpha_2 = \min(-b_{0j}/b'_{0j})$.

4. Среди чисел b'_{0j} есть равные нулю. Пусть, например, $b'_{0s} = 0$. Тогда $b_{0s} + b'_{0s}t = b_{0s} \geqslant 0$, т. е. признак оптимальности выполняется. Следовательно, на равные нулю элементы последней строки можно не обращать внимания.

Таким образом, рассмотрены все случаи, которые могут встретиться при решении задачи, имеющей конечный максимум.

После получения отрезка $[\alpha_1; \alpha_2]$ сравниваем его с данным отрезком $[\alpha; \beta]$. Если $\alpha_2 \geqslant \beta$, то весь отрезок $[\alpha; \beta]$ покрывается отрезком $[\alpha_1; \alpha_2]$ и задача решена: для всех $t \in [\alpha; \beta]$ функция f_t достигает максимума в той же вершине $\mathbf{x}_1^*$ многогранника M, в которой достигается максимум f_α, и $\max f_t = b_{00} + b'_{00}t$.

Если же $\alpha_2 < \beta$, то на отрезке $[\alpha; \alpha_2]$ максимум f_t будет в той же вершине, в которой достигается максимум f_α, а оставшийся отрезок $[\alpha_2; \beta]$ требует дальнейшего исследования. Отрезок $[\alpha_2; \beta]$ можно исследовать так же, как и отрезок $[\alpha; \beta]$, т. е. положить $t = \alpha_2$ и т. д.

Нам остается рассмотреть случай неограниченности целевой функции f_t при $t = \alpha$. При этом в некотором столбце, например с номером s, элемент f_α-строки отрицателен: $b_{0s} + \alpha b'_{0s} < 0$, а все остальные элементы s-го столбца неположительны: $b_{is} \leqslant 0$. Сразу же делать вывод о неограниченности функции f_t при любом t не следует. Из рис. 9.2 видно, что при $t = \alpha'$ функция $f_{\alpha'}$ не ограничена, а при $t = \alpha''$ она достигает максимума в вершине A. В этом случае рассматривается знак числа b'_{0s}. Если $b'_{0s} \leqslant 0$, то при всех $t \geqslant \alpha$ сумма $b_{0s} + tb'_{0s} < 0$ и функция f_t не ограничена для всех $t \in [\alpha; \beta]$. Если же $b'_{0s} > 0$, то для всех $t < -b_{0s}/b'_{0s}$ сумма $b_{0s} + tb'_{0s} < 0$ и функция не ограничена для $t \in [\alpha; -b_{0s}/b'_{0s})$. Что же касается отрезка $[-b_{0s}/b'_{0s}; \beta]$ при $\beta > -b_{0s}/b'_{0s}$, то он подлежит дальнейшему исследованию. Для этого полагаем $t = -b_{0s}/b'_{0s}$ и применяем последовательно рассмотренную процедуру решения задачи. Если же $\beta < -b_{0s}/b'_{0s}$, то решение задачи заканчивается раньше.

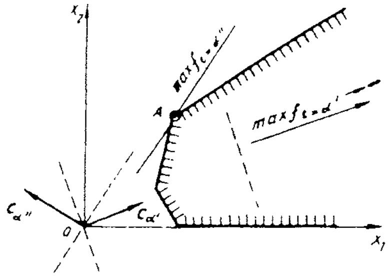


Рис. 9.2

Пример 9.1. Решить ЗЛПП:

$$\max z_t = (1-t)x_1 + x_2, \quad t \in [2; 10];$$

$$\left.\begin{array}{rcrl} x_1 & + & x_2 & \leqslant 7, \\ -5x_1 & + & 2x_2 & \leqslant 1, \\ x_1 & - & x_2 & \leqslant 1, \\ x_1 & & & \geqslant 0, \\ & & x_2 & \geqslant 0. \end{array}\right\}$$

Решение. При $t = 2$ целевая функция принимает вид $z_2 = -x_1 + x_2$. Систему ограничений-неравенств преобразовываем в эквивалентную систему уравнений:

$$\left.\begin{array}{rcrcrcrcrl} x_1 & + & x_2 & + & x_3 & & & & & = 7, \\ x_1 & - & x_2 & & & + & x_4 & + & & = 1, \\ -5x_1 & + & 2x_2 & & & & & + & x_5 & = 1, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,5}).$$

Полученную ЗЛП решаем симплекс-методом (табл. 9.3 — 9.5), приписывая снизу к таблицам по две строки для коэффициентов целевой функции $z_t = (1-t)x_1 + x_2$, что не вносит никаких изменений в процедуру симплекс-метода. Элементы указанных строк преобразовываются в процессе решения по тем же правилам, что и элементы других строк таблицы.

Из табл. 9.5 находим $\max z_2 = 23/7$, достигаемый при $x_1^* = 13/7$, $x_2^* = 36/7$.

Таблица 9.3

БП	1	СП	
		$-x_1$	$-x_2$
$x_3 =$	7	1	1
$x_4 =$	1	1	-1
$x_5 =$	1	-5	$\boxed{2}$
$z_2 =$	0	1	-1
z_t	0	-1	-1
	0	1	0

Таблица 9.4

БП	1	СП	
		$-x_1$	$-x_5$
$x_3 =$	13/2	$\boxed{7/2}$	$-1/2$
$x_4 =$	3/2	$-3/2$	1/2
$x_2 =$	1/2	$-5/2$	1/2
$z_2 =$	1/2	$-3/2$	1/2
z_t	1/2	$-7/2$	1/2
	0	1	0

Таблица 9.5

БП	1	СП	
		$-x_3$	$-x_5$
$x_1 =$	13/7	2/7	$-1/7$
$x_4 =$	30/7	3/7	2/7
$x_2 =$	36/7	5/7	1/7
$z_2 =$	23/7	3/7	2/7
z_t	7	1	0
	$-13/7$	$-2/7$	1/7

Определяем множество всех значений параметра t, при которых $\max z_t$ достигается в полученной вершине $\mathbf{x}_1^* = (13/7;\ 36/7)$. Так как в последней строке табл. 9.5 есть положительные и отрицательные числа, находим:

$$\alpha_1 = \max_{b'_{0j}>0}\left(-\frac{b_{0j}}{b'_{0j}}\right) = \max_{b'_{0j}>0}\left(-\frac{0}{1/7}\right) = 0,$$

$$\alpha_2 = \min_{b'_{0j}<0}\left(-\frac{b_{0j}}{b'_{0j}}\right) = \min_{b'_{0j}<0}\left(\frac{-1}{-2/7}\right) = 7/2,$$

т. е. вершина $\mathbf{x}_1^*$ оптимальна для всех $t \in [0; 7/2]$; при этом $\max z_t = 7 - 13t/7$.

Так как $\alpha_2 = 7/2 < 10 = \beta$, частичный отрезок $[2; 7/2]$ исключаем из данного отрезка $[2; 10]$ и продолжаем решение задачи для оставшегося отрезка $[7/2; 10]$. Полагаем $t = 7/2$ в целевой функции и получаем $z_{7/2} = (1 - 7/2)x_1 + x_2 = -5/2x_1 + x_2$.

По типу табл. 9.3 составляем табл. 9.6 со строкой для функции $z_{7/2}$ и прежней системы ограничений. Затем находим оптимальную вершину $\mathbf{x}_2^*$ (табл. 9.6, 9.7) с координатами $x_1^* = 0,\ x_2^* = 1/2$, в которой функция $z_{7/2}$ достигает $\max z_{7/2} = 1/2$.

Таблица 9.6

БП	1	СП	
		$-x_1$	$-x_2$
$x_3 =$	7	1	1
$x_4 =$	1	1	-1
$x_5 =$	1	-5	$\boxed{2}$
$z_{7/2} =$	0	5/2	-1
z_t	0	-1	-1
	0	1	0

Таблица 9.7

БП	1	СП	
		$-x_1$	$-x_5$
$x_3 =$	13/2		
$x_4 =$	3/2		
$x_2 =$	1/2		
$z_{7/2} =$	1/2	0	1/2
z_t	1/2	$-7/2$	1/2
	0	1	0

Находим все значения t, при которых $\max z_t = 1/2$ достигается в $\mathbf{x}_2^*$: $t \in [7/2; \infty)$.

Итак, отрезок $[2; 10]$ разбивается на два: $[2; 7/2]$ и $[7/2; 10]$. Для $t \in [2; 7/2]$ $\max z_t = 7 - 13t/7$ достигается в вершине $\mathbf{x}_1^* = (13/7; 36/7)$; для $t \in [7/2; 10]$ $\max z_t = 1/2$ достигается в вершине $\mathbf{x}_2^* = (0; 1/2)$.

10. СТОХАСТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

10.1. ПОНЯТИЕ О СТОХАСТИЧЕСКИХ ЗАДАЧАХ И СТОХАСТИЧЕСКОМ ПРОГРАММИРОВАНИИ

Оптимизационные задачи и их модели рассматривались в этой книге в предположении, что вся исходная информация задана строго однозначно. Такие задачи и модели называют *детерминированными.* В действительности же детерминированные модели часто оказываются неадекватными реальным процессам в экономике. Это объясняется неполнотой, неточностью данных, на основе которых формируется модель. В одних случаях некоторые (а возможно, и все) параметры модели носят вероятностный характер. Тогда говорят о ситуациях, связанных с риском. В других случаях имеющаяся информация не позволяет составить представление о характере изменения параметров, описывающих изучаемый процесс. Такие ситуации называют неопределенными. Оптимизационные задачи, при постановке которых нет исчерпывающих данных об их условиях, называют *стохастическими.* Поскольку стохастические задачи приходится ставить и решать при недостаточной информации, это ведет к снижению экономической эффективности получаемых решений.

Для исследования описанных ситуаций разрабатываются специальные методы, объединяемые в разделе математического программирования, называемом *стохастическим программированием.* Стохастическое программирование изучает теорию и методы решения условных экстремальных задач при неполной информации об условиях задач.

Различают пассивное и активное стохастическое программирование. *Пассивное стохастическое программирование* — это совокупность приемов, позволяющих находить наилучшие решения и экстремальные значения целевых функций в оптимизационных задачах со случайными исходными данными. При этом используются, в частности, идеи параметрического программирования. *Активное стохастическое программирование* — это совокупность приемов, позволяющих развивать методы выбора решений в условиях риска и неопределенности.

В стохастическом программировании исследуются одно-, двух- и многоэтапные задачи.

Одноэтапными называют задачи, в которых последовательность поступления исходной информации не имеет значения при выборе решения, оно принимается один раз и в дальнейшем не корректируется. Такие задачи могут порождаться детерминированными оптимизационными задачами, когда исходные данные теряют определенность и приобретают случайный характер. В одноэтапных задачах по-разному выбираются целевая функция и характер ограничений. В качестве целевой функции может рассматриваться вероятность попадания решения в некоторую, вообще говоря, случайную область или математическое ожидание некоторой функции от решения. В одних случаях ограничения задачи могут удовлетворяться при всех возможных значениях случайных параметров (жесткая постановка). В других случаях требуется, чтобы вероятность попадания решения в допустимую область была не меньше данного числа (модель с вероятностными ограничениями). В каждой конкретной задаче приходится специально оговаривать, что называть планом и оптимальным планом.

Естественный на первый взгляд путь замены случайных параметров их средними значениями и поиск оптимальных планов полученных таким образом детерминированных задач во всех случаях не всегда оправдан. При усреднении параметров условий задачи может быть нарушена адекватность модели изучаемому явлению. Оптимальный план детерминированной задачи с усредненными параметрами не всегда удовлетворяет условиям задачи при различных возможных комбинациях параметров ограничений. Этот прием применяют лишь для приближенного решения стохастической задачи.

На практике чаще других встречаются *двухэтапные стохастические задачи*. Подобные задачи возникают, например, при планировании выпуска продукции в случае, когда отсутствуют данные о спросе на нее. В такой ситуации сначала принимается предварительное решение об объеме выпуска на основе имеющейся информации (1-й этап), затем после установления спроса принимается корректирующее решение (2-й этап). При этом предварительное решение не должно исключать возможность его коррекции на втором этапе. Кроме того, предварительный и корректирующий планы согласовывают так, чтобы обеспечивались минимальные средние затраты за оба этапа. Для ряда конкретных постановок двухэтапных стохастических задач разработаны достаточно надежные методы решения.

В *многоэтапных (динамических) задачах* по мере получения информации о развивающемся процессе имеется возможность неоднократно корректировать решение, учитывая исходные ограничения и априорные статистические характеристики случайных параметров, описывающих процесс на каждом этапе. Многоэтапные задачи могут быть как с жесткими, так и с вероятностными ограничениями. Примерами многоэтапных стохастических задач являются задачи перспективного планирования, регулирования технологических процессов, оперативного управления космическими объектами и др. Многоэтапные стохастические задачи представляют для практики наибольший интерес.

10.2. ОДНОШАГОВЫЕ И МНОГОШАГОВЫЕ СТОХАСТИЧЕСКИЕ ЗАДАЧИ ПРОИЗВОДСТВЕННОГО ХАРАКТЕРА

Задача об оптимальном распределении самолетов по маршрутам. В распоряжении авиакомпании находится m различных типов самолетов, предназначенных для обслуживания n маршрутов. Требуется распределить самолеты по маршрутам на некоторый период времени. Может быть использовано a_i самолетов i-го типа. Если самолет i-го типа работает на j-м маршруте, то он может перевезти d_{ij} пассажиров за некоторый промежуток времени (предполагается, что сюда включены путешествия в оба конца). При этом стоимость эксплуатации самолета равна c_{ij}. Допустим, что спрос на билеты одинаков на любом направлении, и если пассажир выходит в промежуточном пункте, то кто-то сразу же занимает его место. Однако спрос v_j на j-м маршруте точно предсказать нельзя, поэтому его следует рассматривать как случайную величину, непрерывно распределенную с плотностью вероятности $\varphi_j(v_j)$.

Обозначим через x_{ij} количество самолетов i-го типа ($i = \overline{1, m}$), обслуживающих j-й маршрут ($j = \overline{1, n}$). Тогда количество y_j пассажиров, которое можно перевезти на всех x_{ij} самолетах j-го маршрута, выразится суммой

$$y_j = \sum_{i=1}^{m} d_{ij} x_{ij} \quad (j = \overline{1, n}). \tag{10.1}$$

Если $y_j < v_j$, т. е. спрос превышает возможности, компания не получит той выручки, которую могла бы получить при наличии достаточного количества мест в самолетах j-го маршрута. В этом случае из-за недостатка мест не будет перевезено $v_j - y_j$ пассажиров и потери будут пропорциональны неудовлетворенному спросу $v_j - y_j$, т. е. $\pi_j(v_j - y_j)$, где π_j — потери за счет каждого пассажира, не получившего место в самолете.

Средние ожидаемые потери от неудовлетворенного спроса на j-м маршруте определяются математическим ожиданием M_j' случайной величины $\pi_j(v_j - y_j)$.

Известно, что математическое ожидание $M(X)$ случайной величины X, непрерывно распределенной на числовой оси с плотностью вероятности $\varphi(x)$, вычисляется по формуле

$$M(X) = \int_{-\infty}^{\infty} x\varphi(x)\,dx.$$

Если $y = f(X)$ — функция случайной величины X, то ее математическое ожидание $M(f(X))$ выражается формулой

$$M(f(X)) = \int_{-\infty}^{\infty} f(x)\varphi(x)\,dx.$$

В соответствии с этим в нашем случае

$$M_j' = \int_{y_j}^{\infty} \pi_j(v_j - y_j)\varphi_j(v_j)\,dv_j \quad (y_j \leqslant v_j < \infty). \tag{10.2}$$

Если же $y_j > v_j$, т. е. количество мест на j-м маршруте превышает спрос, то вновь компания несет убытки, выражающиеся величиной, пропорциональной количеству мест, оставшихся свободными, т. е. $k_j(y_j - v_j)$, где k_j — убыток от каждого незанятого места.

Средние ожидаемые убытки от всех $y_j - v_j$ свободных мест определяются математическим ожиданием M_j'' случайной величины $k_j(y_j - v_j)$:

$$M_j'' = \int_0^{y_j} k_j(y_j - v_j)\varphi_j(v_j)\,dv_j \quad (0 \leqslant v_j < y_j). \tag{10.3}$$

Эксплуатационные расходы, связанные с работой x_{ij} самолетов на j-м маршруте, составляют

$$\sum_{i=1}^{m} c_{ij}x_{ij}. \tag{10.4}$$

Средние ожидаемые суммарные потери от эксплуатации самолетов на всех маршрутах из-за неудовлетворенного спроса и избытка мест выразятся суммой по всем n маршрутам, определяемым выражениями (10.2) — (10.4):

$$z = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} c_{ij}x_{ij} + \int_{y_j}^{\infty} \pi_j(v_j - y_j)\varphi_j(v_j)\,dv_j + \right.$$

$$\left. + \int_0^{y_j} k_j(y_j - v_j)\varphi_j(v_j)\,dv_j\right). \tag{10.5}$$

Найдем такие x_{ij}, которые минимизировали бы функцию (10.5). Заметим, что y_j выражаются через x_{ij} с помощью равенства (10.1).

Значения x_{ij} должны удовлетворять следующим условиям: количество самолетов i-го типа, используемых на всех n маршрутах, не должно превышать общего количества a_i самолетов, находящихся в распоряжении данной авиакомпании, т. е.

$$\sum_{j=1}^{n} x_{ij} \leqslant a_i \quad (i = \overline{1, m}). \tag{10.6}$$

Кроме того, из равенства (10.1) имеем

$$\sum_{i=1}^{m} d_{ij}x_{ij} - y_j = 0 \quad (j = \overline{1, n}). \tag{10.7}$$

По смыслу задачи

$$x_{ij} \geqslant 0 \quad (i = \overline{1,m};\ j = \overline{1,n}), \tag{10.8}$$

$$y_j \geqslant 0 \quad (j = \overline{1,n}). \tag{10.9}$$

Можно показать, что целевая функция (10.5) выпуклая и сепарабельная. Тогда задача решается методом линейной аппроксимации.

Используя равенства (10.1), можно исключить y_j из целевой функции (10.5). Однако если сделать это, целевая функция перестанет быть сепарабельной и решение задачи значительно усложнится. Поэтому приведенная форма (10.5) – (10.9) задачи предпочтительнее.

Можно считать, что рассмотренная одношаговая стохастическая задача порождена линейной детерминированной задачей с определенным значением спроса. Ввиду того что спрос мы считали величиной случайной, целевая функция (10.5) стала нелинейной, а задача — стохастической.

Перейдем к рассмотрению многошаговых стохастических задач.

Задача о наилучшем использовании механизмов. Прокладка тоннелей строящегося метрополитена ведется на участках A и $Б$ с различными горно-геологическими условиями. Если проходческий щит работает на участке A, то с вероятностью p_1 выбирается r_1-я часть всего объема x грунта, подлежащего выемке на этом участке, и с вероятностью $1-p_1$ щит выходит из строя. Если же щит работает на участке $Б$, то с вероятностью p_2 выбирается r_2-я часть всего объема y грунта, подлежащего выемке на этом участке, и с вероятностью $1 - p_2$ щит потребует ремонта. В какой последовательности надлежит использовать щит на участках A и $Б$, чтобы объем выбранного грунта до выхода щита из строя был максимальным?

Решим задачу методом динамического программирования. Разобьем период работы щита на этапы. Проходку можно начать либо на участке A, либо на участке $Б$. Если щит не вышел из строя на предыдущем этапе, то надо решить вопрос, на каком участке его целесообразно использовать на последующем этапе.

Обозначим через $f_n(x, y)$ максимальный ожидаемый объем выбранного грунта за n этапов при условии, что на начало проходческих работ объемы грунта, подлежащие выемке на участках A и $Б$, равны соответственно x и y.

Рассмотрим сначала одноэтапный период. Допустим, что проходческие работы выполняются на участке A; тогда ожидаемый объем выбранного грунта можно оценить величиной

$$p_1 r_1 x; \tag{10.10}$$

если же работы ведутся на участке $Б$, — величиной

$$p_2 r_2 y. \tag{10.11}$$

Оптимальный выбор участка в одноэтапном периоде будет соответствовать наибольшему из значений (10.10) и (10.11). Следовательно,

$$f_1(x,y) = \max\left(p_1 r_1 x, p_2 r_2 y\right). \tag{10.12}$$

Предположим теперь, что за первым следует еще N этапов. Если проходка ведется на участке A, то при безаварийной работе на первом этапе будет выбрано $r_1 x$ единиц грунта и на начало второго этапа на этом участке объем работы составит $(1 - r_1)x$ единиц грунта.

Если допустить, что на оставшихся N этапах щит будет использоваться оптимально, то ожидаемый максимальный объем выбранного грунта на последних N этапах составит $f_N((1-$ $-r_1)x, y)$ единиц. В этом случае общий объем грунта, который может быть выбран за $N+1$ этапов при условии безаварийной работы щита, выразится величиной $r_1 x + f_N((1 - r_1)x, y)$. С учетом же вероятностного характера процесса максимальный ожидаемый объем выбранного грунта будет равен

$$f_{N+1}^{A}(x,y) = p_1(r_1 x + f_N((1 - r_1)x, y)). \tag{10.13}$$

Если работа велась сначала на участке $Б$, то аналогично можно получить

$$f_{N+1}^{Б}(x,y) = p_2(r_2 y + f_N(x, (1 - r_2)y)). \tag{10.14}$$

Оптимальное решение всей задачи определится максимальным ожидаемым объемом выбранного грунта, отвечающим наибольшему из значений (10.13) и (10.14):

$$f_{N+1}(x,y) = \max \left\{ \begin{array}{l} f_{N+1}^{A}(x,y), \\ f_{N+1}^{Б}(x,y) \end{array} \right. =$$

$$= \max \begin{cases} p_1(r_1 x + f_N((1 - r_1)x, y)), \\ p_2(r_2 y + f_N(x, (1 - r_2)y)). \end{cases} \tag{10.15}$$

Это основное функциональное соотношение Беллмана для данной задачи. Используя равенства (10.12) и (10.15), можно последовательно составить оптимальную стратегию применения проходческого щита в $(N + 1)$-этапном периоде.

Задача о наиболее рациональном управлении запасами продукта. На складе хранится запас некоторого продукта в течение n периодов времени. На начало первого периода запас составляет y_0 единиц. Количество продукта, отправляемого со склада в j-й период, определяется спросом v_j на этот продукт. При этом спрос v_j в каждый данный период считается случайной величиной с плотностью вероятности $\varphi_j(v_j)$. Предполагается, что для различных периодов эти случайные величины независимы. Стоимость доставки единицы заказанного продукта с предприятия на склад в j-й период равна c_j (заказ делается в начале каждого периода, а доставка его на склад производится до конца этого же периода). Стоимость хранения единицы продукта в j-м периоде равна k_j, а стоимость хранения запаса пропорциональна его количеству y_j в конце периода и составляет $k_j y_j$. За отсутствие продукта на складе при наличии спроса на него в j-м периоде начисляется штраф, равный π_j, в расчете на единицу продукта. Штраф увеличивается пропорционально количеству отказов s_j к концу j-го периода и равняется $\pi_j s_j$.

Требуется так спланировать поступление продукта на склад, чтобы минимизировались суммарные расходы, связанные с его доставкой, хранением и штрафами за отсутствие продукта при наличии спроса.

Обозначим через x_j количество единиц продукта, заказываемого на j-й период. Тогда к концу j-го периода запас продукта

$$y_j = y_0 + \sum_{i=1}^{j} x_i - \sum_{i=1}^{j} v_i. \tag{10.16}$$

Очевидно, что работа склада должна строиться таким образом, чтобы поддерживать $y_j \geqslant 0$. Однако это не всегда удается, и число отказов из-за отсутствия продукта на складе к концу j-го периода оказывается равным

$$s_j = \sum_{i=1}^{j} v_i - \sum_{i=1}^{j} x_i - y_0 \quad (s_j \geqslant 0). \qquad (10.17)$$

Наличие запаса и штрафа — взаимоисключающие моменты: если в течение какого-то j-го периода $y_j \geqslant 0$, то $s_j = 0$, и наоборот, т. е. если приходится платить за хранение в j-й период, то штрафа за отказ в этот период не будет. Оба описанных момента можно отразить посредством функции

$$F_j\left(y_0 + \sum_{i=1}^{j} x_i - \sum_{i=1}^{j} v_i\right) = \begin{cases} k_j y_j, & \text{если } y_j \geqslant 0, \\ \pi_j s_j, & \text{если } s_j > 0, \end{cases} \qquad (10.18)$$

где y_j, s_j определяются выражениями (10.16) и (10.17) соответственно.

Суммарные расходы склада за n периодов с учетом равенств (10.18) можно записать в виде

$$R = \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} F_j\left(y_0 + \sum_{i=1}^{j} x_i - \sum_{i=1}^{j} v_i\right). \qquad (10.19)$$

Поскольку спрос v_j — случайная величина, выражение (10.19) является функцией случайной величины v_j. Найдем математическое ожидание $M(R) = \overline{R}$.

Напомним, что если $X_1, \ldots, X_n$ — система непрерывных случайных величин с плотностью вероятности $\Phi(x_1, \ldots, x_n)$, а $Y = f(X_1, \ldots, X_n)$ — функция этих случайных величин, то математическое ожидание величины Y

$$M(Y) = M(f(X_1, \ldots, X_n)) =$$

$$= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\Phi(x_1, \ldots, x_n)\, dx_1 \ldots dx_n. \qquad (10.20)$$

Если непрерывные случайные величины $X_1, \ldots, X_n$ независимо распределены с плотностями вероятности $\varphi_1(x_1), \ldots, \varphi_n(x_n)$, то плотность вероятности $\Phi(x_1, \ldots, x_n)$ появления некоторой определенной комбинации величин $X_1, \ldots, X_n$ равна произведению плотностей вероятности появления каждой из величин, т. е.

$$\Phi(x_1, \ldots, x_n) = \varphi_1(x_1) \ldots \varphi_n(x_n) = \prod_{j=1}^{n} \varphi_j(x_j). \qquad (10.21)$$

С учетом равенства (10.21) формула (10.20) примет вид

$$M(Y) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \prod_{j=1}^{n} \varphi_j(x_j)\, dx_1 \ldots dx_n. \quad (10.22)$$

Как видно из равенства (10.19), в нашем случае R есть функция n независимых непрерывно распределенных на интервале $(0; \infty)$ с плотностями вероятности $\varphi_1(v_1), \ldots, \varphi_n(v_n)$ случайных величин $v_1, \ldots, v_n$. Поэтому в соответствии с формулой (10.22) математическое ожидание $M(R) = \overline{R}$ функции R

$$\overline{R} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \left(\sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} F_j \left(y_0 + \sum_{i=1}^{j} x_i - \right.\right.$$

$$\left.\left. - \sum_{i=1}^{j} v_i \right)\right) \prod_{j=1}^{n} \varphi_j(v_j)\, dv_1 \ldots dv_n. \qquad (10.23)$$

Функция (10.23) выражает математическое ожидание расходов склада за n периодов. Задача состоит в выборе $x_j \geqslant 0$, минимизирующих $\overline{R}$.

Сразу выбрать все n значений x_j нельзя. В самом деле, решение о том, какое количество продукта следует заказать в начале, например, второго периода, зависит от того, какое количество продукта было завезено на склад и каков был спрос в течение этого периода, т. е. оптимальное значение x_2^* является функцией y_2 — запаса в начале второго периода. Аналогично дело обстоит для любого k-го периода, т. е. $x_k^* = x_k^*(y_k)$. Такой подход к решению соответствует концепции метода динамического программирования: решение на каждом шаге (периоде) выбирается исходя из результатов предыдущего шага (периода).

ПРИЛОЖЕНИЕ

Использование обратной матрицы в экономико-математическом анализе

Обратная матрица

Среди обычных чисел существует особенное число 1, дающее в произведении с любым числом a снова a, т. е. $a \cdot 1 = 1 \cdot a = a$. Оказывается, что среди квадратных матриц аналогичную роль играет единичная матрица

$$E = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Пусть дана квадратная матрица A n-го порядка:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

Легко проверить, что $AE = EA = A$.

Известно, что в арифметике чисел для каждого числа $a \neq 0$ существует обратное число $1/a = a^{-1}$, дающее в произведении с a единицу, т. е. $aa^{-1} = a^{-1}a = 1$.

В матричной алгебре существуют матрицы, обладающие аналогичным свойством, причем роль условия $a \neq 0$ играет условие $\Delta \neq 0$, где Δ — определитель матрицы A:

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Матрица, определитель которой отличен от нуля, называется *невырожденной (неособенной)*, а матрица с равным нулю определителем — *вырожденной (особенной)*.

Матрица A^{-1} называется *обратной* для матрицы A, если

$$AA^{-1} = A^{-1}A = E. \quad (1)$$

Имеет место следующая важная теорема, которую мы приводим без доказательств: *если матрица A невырожденная, то существует, и притом единственная, обратная матрица A^{-1}, для которой одновременно выполняются равенства $A^{-1}A = E$ и $AA^{-1} = E$.*

Доказательство этой теоремы носит конструктивный характер, и в ходе его получается следующая формула для нахождения обратной матрицы:

$$A^{-1} = \frac{1}{\Delta}\widetilde{A}, \tag{2}$$

где Δ — определитель данной матрицы A; $\widetilde{A}$ — присоединенная к A матрица, т. е. матрица, составленная из алгебраических дополнений A_{ij} элементов a_{ij} матрицы A и имеющая вид

$$\widetilde{A} = \begin{bmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{bmatrix}.$$

Пример 1. Для матрицы

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

найти обратную матрицу A^{-1}, пользуясь формулой (2).

Решение. Вычисляем определитель матрицы A:

$$\Delta = \begin{vmatrix} 1 & -1 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 2 \end{vmatrix} = 5.$$

Поскольку $\Delta \neq 0$, то матрица A невырожденная, а потому для нее существует обратная матрица A^{-1}.

Составляем присоединенную матрицу:

$$\widetilde{A} = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}.$$

Вычисляем алгебраические дополнения A_{ij}, пользуясь формулой

$$A_{ij} = (-1)^{i+j} M_{ij},$$

где M_{ij} — минор элемента a_{ij} определителя Δ, т. е. определитель $(n-1)$-го порядка, полученный из определителя Δ вычеркиванием i-й строки и j-го столбца, на пересечении которых находится элемент a_{ij}. Так,

$$A_{11} = (-1)^{1+1} \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1.$$

Аналогично находим: $A_{21} = 3$, $A_{31} = -2$, $A_{12} = -3$, $A_{22} = 1$, $A_{32} = 1$, $A_{13} = 1$, $A_{23} = -2$, $A_{33} = 3$.

Присоединенная матрица принимает следующий вид:

$$\widetilde{A} = \begin{bmatrix} 1 & 3 & -2 \\ -3 & 1 & 1 \\ 1 & -2 & 3 \end{bmatrix}.$$

В соответствии с формулой (2) обратная матрица имеет вид

$$A^{-1} = \begin{bmatrix} 1/5 & 3/5 & -2/5 \\ -3/5 & 1/5 & 1/5 \\ 1/5 & -2/5 & 3/5 \end{bmatrix}.$$

Из примера видно, что вычисление присоединенной матрицы очень трудоемко, особенно для матриц высокого порядка.

Рассмотрим другой способ численного определения обратной матрицы, в котором используются алгебраические преобразования, составляющие вычислительный аппарат, широко применяемый в математическом программировании. Речь идет о полных гауссовых, а также жордановых исключениях. В связи с этим и способ назовем *способом исключений*.

Обращение матрицы способом исключений

Пусть A — квадратная невырожденная матрица n-го порядка, обратную для которой необходимо найти.

С этой целью рассмотрим систему n линейных уравнений с n неизвестными $x_1, x_2, \ldots, x_n$, коэффициентами при которых являются элементы рассматриваемой матрицы A, а свободными членами — пока не определенные числа $y_1, y_2, \ldots, y_n$, т. е. систему вида

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = y_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = y_n, \end{array}\right\} \tag{3}$$

или в матричной форме

$$AX = Y, \tag{4}$$

где X — матрица-столбец $[x_1 \ x_2 \ \ldots \ x_n]^T$ неизвестных; Y — матрица-столбец $[y_1 \ y_2 \ \ldots \ y_n]^T$ свободных членов.

Поскольку матрица A невырожденная, то ее определитель отличен от нуля. Это означает, что система (3) (или, что то же, (4)) имеет единственное решение. Получим его в матричной форме. Для этого умножим обе части равенства (4) на обратную матрицу A^{-1} (она существует, поскольку A — невырожденная матрица):

$$A^{-1}AX = A^{-1}Y.$$

С учетом равенства (1) запишем полученное равенство в виде $EX = A^{-1}Y$ или

$$X = A^{-1}Y. \tag{5}$$

Обозначим элементы искомой матрицы A^{-1} через b_{ij}. Тогда

$$A^{-1} = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix}.$$

Наша задача — найти элементы b_{ij} обратной матрицы A^{-1}.

Для большей наглядности будем сопровождать теоретические рассуждения конкретным числовым примером. Предположим, что для матрицы

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 2 \\ 4 & 1 & 4 \end{bmatrix}$$

необходимо найти обратную матрицу A^{-1}. Устанавливаем, что определитель Δ этой матрицы равен 6, т. е. отличен от нуля. Значит, обратная матрица существует. Система (3) в данном случае запишется в виде

$$\left.\begin{aligned} x_1 + x_2 + 2x_3 &= y_1, \\ 2x_1 - x_2 + 2x_3 &= y_2, \\ 4x_1 + x_2 + 4x_3 &= y_3, \end{aligned}\right\} \tag{6}$$

а искомая обратная матрица — в виде

$$A^{-1} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}.$$

Вернемся к теории вопроса. С целью определения элементов b_{ij} обратной матрицы A^{-1} положим в равенстве (4) $Y = E_1 = [1\ \ 0\ \ \ldots\ \ 0]^T$. Тогда получим систему уравнений

$$AX = E_1$$

с вполне определенными свободными членами. Ее решение в соответствии с равенством (5) примет вид

$$X = A^{-1}E_1$$

или

$$\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{bmatrix}.$$

Выполняя справа операцию умножения матриц, получаем

$$\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{11} \\ b_{21} \\ \ldots \\ b_{n1} \end{bmatrix}, \tag{7}$$

откуда $b_{11} = x_1$, $b_{21} = x_2, \ldots, b_{n1} = x_n$. Таким образом, первый столбец элементов обратной матрицы A^{-1} является решением системы уравнений (3), когда свободными членами в ней служат элементы единичной матрицы $E_1 = [1\ \ 0\ \ \ldots\ \ 0]^T$.

Проиллюстрируем это положение на нашем примере. Система (6) с учетом матрицы свободных членов $E_1 = [1\ \ 0\ \ 0]^T$ приобретает вид

$$\left.\begin{aligned} x_1 + x_2 + 2x_3 &= 1, \\ 2x_1 - x_2 + 2x_3 &= 0, \\ 4x_1 + x_2 + 4x_3 &= 0. \end{aligned}\right\} \tag{8}$$

Решим систему (8) методом полного исключения неизвестных. Напомним, что при матричной записи решение этим методом выполняется в следующем порядке. *Выбирается разрешающий элемент (например, на главной диагонали матрицы) и расширенная матрица системы преобразуется по следующим правилам:*

1) разрешающий элемент заменяется единицей;

2) остальные элементы разрешающего столбца заменяются нулями;

3) остальные элементы разрешающей строки делятся на разрешающий элемент;

4) все прочие элементы матрицы вычисляются по правилу прямоугольника: преобразованный элемент равен разности произведений элементов главной и побочной диагоналей, деленной на разрешающий элемент.

В нашем случае последовательно получаем матрицы:

$$\left[\begin{array}{rrr|r} \boxed{1} & 1 & 2 & 1 \\ 2 & -1 & 2 & 0 \\ 4 & 1 & 4 & 0 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 1 & 2 & 1 \\ 0 & \boxed{-3} & -2 & -2 \\ 0 & -3 & -4 & -4 \end{array}\right] \sim$$

$$\sim \left[\begin{array}{rrr|r} 1 & 0 & 4/3 & 1/3 \\ 0 & 1 & 2/3 & 2/3 \\ 0 & 0 & \boxed{-2} & -2 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]. \qquad (8')$$

Из последней матрицы с учетом равенства (7) находим:

$$x_1 = b_{11} = -1, \quad x_2 = b_{21} = 0, \quad x_3 = b_{31} = 1.$$

Итак, первый столбец обратной матрицы найден.

Но далее естественно продолжить процесс определения элементов обратной матрицы аналогично, т. е. положить в равенстве (4) $Y = E_2 = [0 \ 1 \ 0 \ \ldots \ 0]^T$, и тогда после решения уравнения

$$X = A^{-1}E_2$$

или

$$\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \ldots \\ 0 \end{bmatrix}$$

найдем

$$\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} b_{12} \\ b_{22} \\ \ldots \\ b_{n2} \end{bmatrix},$$

откуда $b_{12} = x_1$, $b_{22} = x_2, \ldots, b_{n2} = x_n$, т. е. элементы второго столбца обратной матрицы A^{-1} являются решением системы уравнений (3), когда свободными членами ее служат соответствующие элементы единичной матрицы $E_2 = [0 \ 1 \ 0 \ \ldots \ 0]^T$.

В нашем числовом примере соответствующая система (6) с вектор-столбцом свободных членов $E_2 = [0 \ \ 1 \ \ 0]^T$ имеет следующий вид:

$$\left.\begin{aligned} x_1 + x_2 + 2x_3 &= 0, \\ 2x_1 - x_2 + 2x_3 &= 1, \\ 4x_1 + x_2 + 4x_3 &= 0, \end{aligned}\right\} \tag{9}$$

а ее решение получено в результате тех же самых преобразований, что и в случае системы (8) (системы различаются только свободными членами!):

$$\left[\begin{array}{rrr|r} \boxed{1} & 1 & 2 & 0 \\ 2 & -1 & 2 & 1 \\ 4 & 1 & 4 & 0 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 1 & 2 & 0 \\ 0 & \boxed{-3} & -2 & 1 \\ 0 & -3 & -4 & 0 \end{array}\right] \sim$$

$$\sim \left[\begin{array}{rrr|r} 1 & 0 & 4/3 & 1/3 \\ 0 & 1 & 2/3 & -1/3 \\ 0 & 0 & \boxed{-2} & -1 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 0 & 0 & -1/3 \\ 0 & 1 & 0 & -2/3 \\ 0 & 0 & 1 & 1/2 \end{array}\right], \tag{9'}$$

откуда $x_1 = b_{12} = -1/3$, $x_2 = b_{22} = -2/3$, $x_3 = b_{32} = 1/2$. Таким образом, найден второй столбец обратной матрицы A^{-1}. Описанным путем можно найти и остальные столбцы матрицы A^{-1}.

В рассматриваемом числовом примере придется решить еще одну (третью!) систему типа (6) с вектор-столбцом свободных членов $E_3 = [0 \ \ 0 \ \ 1]^T$:

$$\left.\begin{aligned} x_1 + x_2 + 2x_3 &= 0, \\ 2x_1 - x_2 + 2x_3 &= 0, \\ 4x_1 + x_2 + 4x_3 &= 1. \end{aligned}\right\} \tag{10}$$

В результате уже дважды выполнявшихся преобразований

$$\left[\begin{array}{rrr|r} \boxed{1} & 1 & 2 & 0 \\ 2 & -1 & 2 & 0 \\ 4 & 1 & 4 & 1 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 1 & 2 & 0 \\ 0 & \boxed{-3} & -2 & 0 \\ 0 & -3 & -4 & 1 \end{array}\right] \sim$$

$$\sim \left[\begin{array}{rrr|r} 1 & 0 & 4/3 & 0 \\ 0 & 1 & 2/3 & 0 \\ 0 & 0 & \boxed{-2} & 1 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 0 & 0 & 2/3 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & -1/2 \end{array}\right] \tag{10'}$$

найдем элементы последнего (третьего) столбца обратной матрицы A^{-1}: $x_1 = b_{13} = 2/3$, $x_2 = b_{23} = 1/3$, $x_3 = b_{33} = -1/2$.

Остается из полученных элементов составить обратную матрицу:

$$A^{-1} = \begin{bmatrix} -1 & -1/3 & 2/3 \\ 0 & -2/3 & 1/3 \\ 1 & 1/2 & -1/2 \end{bmatrix}.$$

Приведенные рассуждения позволяют сделать следующий общий вывод: *для нахождения обратной матрицы A^{-1} необходимо решить n систем линейных уравнений вида (3), отличающихся друг от друга лишь столбцами свободных членов, в качестве которых последовательно берутся элементы единичных матриц-столбцов $E_1, E_2, \ldots, E_n$.*

В рассмотренном примере были решены три системы: (8), (9) и (10).

Анализируя рассмотренную методику решения поставленной задачи, можно сделать следующее практически важное заключение: *поскольку системы уравнений вида (3) отличаются друг от друга только свободными членами, то вместо раздельного их решения целесообразно решать все их одновременно, записав в исходную расширенную матрицу вместо одного сразу все* n *столбцов свободных членов, соответствующих элементам матриц-столбцов* $E_1, E_2, \ldots, E_n$, *т. е. в следующем виде:*

$$\left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \ldots & a_{1n} & 1 & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & a_{2n} & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} & 0 & 0 & \ldots & 1 \end{array}\right]; \qquad (11)$$

после этого выполняются полные исключения последовательно с разрешающими элементами $a_{11}, a'_{22}, a''_{33}, \ldots$ *В результате будет получена матрица вида*

$$\left[\begin{array}{cccc|cccc} 1 & 0 & \ldots & 0 & b_{11} & b_{12} & \ldots & b_{1n} \\ 0 & 1 & \ldots & 0 & b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & b_{n1} & b_{n2} & \ldots & b_{nn} \end{array}\right], \qquad (12)$$

в которой справа от вертикальной черты будет находиться искомая обратная матрица

$$A^{-1} = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix}.$$

Этот вывод напрашивается сам собой, если сравнить цепочки (8′), (9′) и (10′) последовательных преобразований расширенных матриц при решении систем (8), (9) и (10), в которых слева от вертикальной черты выполнялись одни и те же преобразования с одними и теми же числовыми данными.

Учитывая все вышесказанное, решение рассмотренного примера целесообразно записать в следующем компактном виде:

$$\left[\begin{array}{rrr|rrr} \boxed{1} & 1 & 2 & 1 & 0 & 0 \\ 2 & -1 & 2 & 0 & 1 & 0 \\ 4 & 1 & 4 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{rrr|rrr} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{-3} & -2 & -2 & 1 & 0 \\ 0 & -3 & -4 & -4 & 0 & 1 \end{array}\right] \sim$$

$$\sim \left[\begin{array}{rrr|rrr} 1 & 0 & 4/3 & 1/3 & 1/3 & 0 \\ 0 & 1 & 2/3 & 2/3 & -1/3 & 0 \\ 0 & 0 & \boxed{-2} & -2 & -1 & 1 \end{array}\right] \sim \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -1 & -1/3 & 2/3 \\ 0 & 1 & 0 & 0 & -2/3 & 1/3 \\ 0 & 0 & 1 & 1 & 1/2 & -1/2 \end{array}\right].$$

Из последней матрицы и выписывается искомая обратная матрица A^{-1}.

Если рассматривать систему всех вектор-столбцов матрицы (11), то понятно, что подсистема единичных вектор-столбцов, стоящих справа от вертикальной черты, образует один из базисов этой системы (назовем его начальным).

Сравнивая матрицы (11) и (12), видим, что в результате последовательных исключений вектор-столбцы данной матрицы A преобразуются в единичные вектор-столбцы, а вектор-столбцы искомой обратной матрицы A^{-1} появляются на месте единичных вектор-столбцов начального базиса; иначе говоря: данная матрица A преобразуется в единичную матрицу, а единичная матрица, стоящая справа от вертикальной черты, — в обратную матрицу A^{-1}. При этом предполагается, что разрешающие элементы выбираются в процессе исключений на главной диагонали матрицы и в последовательности $a_{11}, a'_{22}, a''_{33}, \ldots$ Если разрешающие элементы выбираются в ином порядке (как это и бывает чаще всего в приложениях), то слева от вертикальной черты в заключительной матрице единичные столбцы не будут образовывать единичную матрицу, а значит, матрица справа от черты не будет обратной для матрицы A. В подобной ситуации необходимо так переставить строки заключительной матрицы, чтобы упомянутые единичные столбцы образовали единичную матрицу. Только после этого справа от вертикальной черты появится искомая обратная матрица A^{-1}.

Описанная ситуация почти всегда встречается, например, в математическом программировании, когда приходится выписывать столбцы обратной матрицы из таблицы, полученной в результате последовательных итераций симплексного метода с разрешающими элементами, выбиравшимися по специальным правилам, совсем не гарантирующим сохранения "диагональной" последовательности $a_{11}, a'_{22}, a''_{33}, \ldots$

Ввиду важности сделанного замечания для дальнейшего изложения проиллюстрируем его числовым примером. Вновь обратимся к рассмотренному примеру, но теперь исключения выполним с другими разрешающими элементами: первое исключение, например, — с разрешающим элементом a_{32}, второе — с a'_{13}, третье — с a''_{21}. Тогда последовательно получим матрицы:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 2 & -1 & 2 & 0 & 1 & 0 \\ 4 & \boxed{1} & 4 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} -3 & 0 & \boxed{-2} & 1 & 0 & -1 \\ 6 & 0 & 6 & 0 & 1 & 1 \\ 4 & 1 & 4 & 0 & 0 & 1 \end{array}\right] \sim$$

$$\sim \left[\begin{array}{ccc|ccc} 3/2 & 0 & 1 & -1/2 & 0 & 1/2 \\ \boxed{-3} & 0 & 0 & 3 & 1 & -2 \\ -2 & 1 & 0 & 2 & 0 & -1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & 1/2 & -1/2 \\ 1 & 0 & 0 & -1 & -1/3 & 2/3 \\ 0 & 1 & 0 & 0 & -2/3 & 1/3 \end{array}\right].$$

Справа от вертикальной черты в последней матрице вновь содержатся элементы обратной матрицы A^{-1}. Однако, прежде чем сформировать из них обратную матрицу, необходимо поменять местами сначала вторую и первую строки:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1/3 & 2/3 \\ 0 & 0 & 1 & 1 & 1/2 & -1/2 \\ 0 & 1 & 0 & 0 & -2/3 & 1/3 \end{array}\right],$$

а затем третью и вторую:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -1/3 & 2/3 \\ 0 & 1 & 0 & 0 & -2/3 & 1/3 \\ 0 & 0 & 1 & 1 & 1/2 & -1/2 \end{array}\right].$$

Вот теперь подматрица, расположенная справа от вертикальной черты, представляет собой искомую обратную матрицу A^{-1}.

Завершающие преобразования, выполненные в этом примере, будут встречаться практически во всех последующих примерах.

Расчет пределов устойчивости двойственных оценок

Задача нахождения обратной матрицы возникает в линейном программировании, например в ходе экономико-математического анализа исследуемого экономического процесса, когда необходимо знать нижние и верхние границы устойчивости двойственных оценок при изменении ограничений. Оказывается, для вычисления пределов устойчивости требуется матрица, обратная для матрицы базиса оптимального плана.

Обратимся к конкретному примеру. Рассмотрим ЗЛП:

$$\max Z = 11x_1 + 6x_2 + 9x_3 + 6x_4;$$

$$\left.\begin{aligned} 2x_1 + x_2 + 3x_3 \qquad &\leqslant 20,\\ 3x_1 + x_2 + x_3 + 2x_4 &\leqslant 37,\\ x_2 + x_3 + 4x_4 &\leqslant 30, \end{aligned}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,4}).$$

В канонической форме задача принимает следующий вид:

$$\max Z = 11x_1 + 6x_2 + 9x_3 + 6x_4 + 0x_5 + 0x_6 + 0x_7;$$

$$\left.\begin{aligned} 2x_1 + x_2 + 3x_3 + \qquad + x_5 \qquad\qquad &= 20,\\ 3x_1 + x_2 + x_3 + 2x_4 \qquad + x_6 \qquad &= 37,\\ x_2 + x_3 + 4x_4 \qquad\qquad + x_7 &= 30, \end{aligned}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,7}).$$

Дополнительные переменные x_5, x_6 и x_7 образуют начальный базис системы переменных задачи, а коэффициенты при них в ограничениях задачи — единичные вектор-столбцы начального базиса системы вектор-коэффициентов.

В табл. 1 приведено полное решение задачи. Для нахождения оптимального плана $\mathbf{x}^* = (6; 8; 0; 5,5; 0; 0; 0)$, при котором $Z_{\max} = 147$, потребовалось выполнить три итерации симплексного метода. Как видно из табл. 1, базис системы вектор-коэффициентов, в котором план оказался оптимальным, состоит из вектор-коэффициентов при переменных x_1, x_2 и x_4 (см. итерацию 3).

Матрица A, обратная для которой используется при вычислении пределов устойчивости двойственных оценок, составляется из коэффициентов при переменных, образовавших базис оптимального плана. (В нашем случае — из коэффициентов при x_1, x_2 и x_4.) Эти коэффициенты можно выписать либо из системы ограничений задачи, либо непосредственно из табл. 1, где они записаны столбцами под переменными x_1, x_2 и x_4 (итерация 0). В данном случае матрица базиса оптимального плана имеет вид

$$A = \begin{bmatrix} 2 & 1 & 0\\ 3 & 1 & 2\\ 0 & 1 & 4 \end{bmatrix}.$$

Таблица 1

Номер итерации	БП	1	x_1	x_2	x_3	x_4	x_5	x_6	x_7
0	x_5	20	$\boxed{2}$	1	3	0	1	0	0
	x_6	37	3	1	1	2	0	1	0
	x_7	30	0	1	1	4	0	0	1
	Z	0	−11	−6	−9	−6	0	0	0
1	x_1	10	1	0,5	1,5	0	0,5	0	0
	x_6	7	0	−1,5	−3,5	$\boxed{2}$	−1,5	1	0
	x_7	30	0	1	1	4	0	0	1
	Z	110	0	−0,5	7,5	−6	5,5	0	0
2	x_1	10	1	0,5	1,5	0	0,5	0	0
	x_4	3,5	0	−0,25	−1,75	1	−0,75	0,5	0
	x_7	16	0	$\boxed{2}$	8	0	3	−2	1
	Z	131	0	−2	−3	0	1	3	0
3	x_1	6	1	0	−0,5	0	−0,25	0,5	−0,25
	x_4	5,5	0	0	−0,75	1	−0,375	0,25	0,125
	x_2	8	0	1	4	0	1,5	−1	0,5
	Z	147	0	0	5	0	4	1	1

В соответствии с приведенными выше рассуждениями мы утверждаем, что элементы b_{ij} искомой обратной матрицы A^{-1} содержатся в той же табл. 1 (см. итерацию 3), поскольку симплексные преобразования, выполнявшиеся при поиске оптимального плана, представляют собой последовательные полные исключения, т. е. те же самые исключения, которыми мы пользовались ранее для нахождения обратной матрицы.

Процесс образования обратной матрицы A^{-1} начался с первой же итерации симплексного метода. Вычленим из табл. 1 (итерация 0) только те столбцы, которые участвуют в решении этой задачи, и составим из них расширенную матрицу вида (11). Левую подматрицу образуют столбцы x_1, x_2 и x_4, составляющие матрицу A, а правую подматрицу — столбцы x_5, x_6 и x_7, соответствующие начальному единичному базису и записанные так, чтобы они составляли единичную матрицу. Таким образом, получаем матрицу

$$\begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_4 & x_5 & x_6 & x_7 \end{array} \\ \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 1 & 0 & 0 \\ 3 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 4 & 0 & 0 & 1 \end{array}\right] \end{array}$$

Завершается процесс формирования строк обратной матрицы A^{-1}

третьей итерацией симплексного метода, когда столбцы x_1, x_2 и x_4 матрицы A преобразовались в единичные столбцы, а единичные столбцы x_5, x_6 и x_7 — в столбцы элементов обратной матрицы A^{-1}. Выписывая из табл. 1 (итерация 3) указанные столбцы, приходим к матрице

$$\begin{array}{cccccc} x_1 & x_2 & x_4 & x_5 & x_6 & x_7 \end{array}$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -0,25 & 0,5 & -0,25 \\ 0 & 0 & 1 & -0,375 & 0,25 & 0,125 \\ 0 & 1 & 0 & 1,5 & -1 & 0,5 \end{array}\right]. \tag{13}$$

Для того чтобы справа от вертикальной черты получить искомую обратную матрицу, остается переставить вторую и третью строки, так как только после этого слева от черты образуется единичная подматрица, а вся матрица приобретет вид (12).

Итак, после указанной перестановки строк получаем

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -0,25 & 0,5 & -0,25 \\ 0 & 1 & 0 & 1,5 & -1 & 0,5 \\ 0 & 0 & 1 & -0,375 & 0,25 & 0,125 \end{array}\right],$$

откуда

$$A^{-1} = \begin{bmatrix} -0,25 & 0,5 & -0,25 \\ 1,5 & -1 & 0,5 \\ -0,375 & 0,25 & 0,125 \end{bmatrix}.$$

Читателю будет интересно сравнить рассмотренную методику вычленения обратной матрицы из симплексной таблицы с нахождением обратной матрицы методом исключений.

Предположим теперь, что решение рассматриваемой ЗЛП выполнялось с использованием жордановых исключений. Мы приведем лишь начальную (табл. 2) и конечную (табл. 3) жордановы таблицы.

Таблица 2

БП	1	СП			
		$-x_1$	$-x_2$	$-x_3$	$-x_4$
$x_5 =$	20	2	1	3	0
$x_6 =$	37	3	1	1	2
$x_7 =$	30	0	1	1	4
$Z =$	0	−11	−6	−9	−6

Таблица 3

БП	1	СП			
		$-x_5$	$-x_7$	$-x_3$	$-x_6$
$x_1 =$	6	−0,25	−0,25	−0,5	0,5
$x_4 =$	5,5	−0,375	0,125	−0,75	0,25
$x_2 =$	8	1,5	0,5	4	−1
$Z =$	147	4	1	5	1

Здесь есть своя специфика, обусловленная тем, что в жордановы таблицы не включают единичные столбцы, соответствующие базисным переменным, записываемым в столбце "БП". Однако их легко реконструировать, если учесть порядок следования переменных в столбце "БП".

Составим расширенную матрицу (13), соответствующую завершающему шагу жордановых исключений, приведшему к оптимальному плану. К этому моменту закончилось формирование строк искомой обратной матрицы A^{-1}. Всю необходимую информацию мы будем брать только из табл. 2 и 3.

Итак, первые три столбца матрицы (13) — это результат преобразования столбцов x_1, x_2 и x_4, составляющих матрицу A, в единичные столбцы. В данном случае упомянутые единичные столбцы явно в табл. 3 не записаны, так как они соответствуют базисным переменным x_1, x_4 и x_2. Восстанавливая эти столбцы, записываем их в порядке следования столбцов матрицы A, в нашем случае — в порядке x_1, x_2, x_4. Положение единицы в каждом таком столбце определяется номером строки, в которой стоит соответствующая базисная переменная в столбце "БП" последней жордановой таблицы. В нашем случае единица в столбце x_1 стоит на первом месте, так как переменная x_1 находится в первой строке табл. 3, единица в столбце x_2 — на третьем месте, поскольку переменная x_2 находится в третьей строке, единица в столбце x_4 — на втором месте, так как переменная x_4 находится во второй строке.

Что же касается трех других столбцов матрицы (13), которые являются результатом преобразования единичных столбцов x_5, x_6 и x_7 начального базиса, то они в табл. 3 присутствуют и переписываются в матрицу (13) в том же порядке, в каком шли единичные столбцы в начальном базисе, т. е. в порядке x_5, x_6, x_7.

Таким образом, получаем

$$\begin{array}{cccccc} x_1 & x_2 & x_4 & x_5 & x_6 & x_7 \end{array}$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -0,25 & 0,5 & -0,25 \\ 0 & 0 & 1 & -0,375 & 0,25 & 0,125 \\ 0 & 1 & 0 & 1,5 & -1 & 0,5 \end{array}\right].$$

Эта матрица, составленная по данным табл. 2 и 3, тождественна матрице (13), построенной по табл. 1, что вполне естественно, поскольку и в том, и в другом случае решается одна и та же ЗЛП. По последней матрице моментально находится обратная матрица A^{-1}, о чем речь шла выше.

Рассмотренный пример позволяет предложить следующий порядок нахождения матрицы A^{-1}, обратной для матрицы A базиса оптимального плана.

1. *Из конечной таблицы, содержащей оптимальный план ЗЛП, выписать единичные столбцы, в которые преобразовались столбцы матрицы A.*

2. *К полученной матрице приписать после вертикальной черты столбцы конечной таблицы, в которые преобразовались столбцы единичной матрицы, соответствующей начальному базису.*

3. *Переставить в случае необходимости строки сформированной матрицы так, чтобы слева от вертикальной черты образовалась единичная подматрица. При этом справа от черты получается искомая обратная матрица A^{-1}.*

4. *Выписать обратную матрицу A^{-1}.*

Предлагаемая методика может использоваться независимо от того, в какой форме задана ЗЛП и какими таблицами (полными или жордановыми) пользуются при симплексном методе ее решения. Неважно также, какие переменные образуют начальный базис: основные, дополнительные, искусственные или любая их комбинация.

Проиллюстрируем это несколькими примерами.

Пример 2. В табл. 4 приведено решение ЗЛП

$$\max Z = 4x_1 + 3x_2 + 6x_3 + 7x_4;$$

$$\left.\begin{array}{l} 2x_1 + x_2 + x_3 + x_4 \leqslant 280, \\ x_1 + x_3 + x_4 \leqslant 80, \\ x_1 + 2x_2 + x_3 \leqslant 250, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,4}),$$

каноническая форма которой имеет вид

$$\max Z = 4x_1 + 3x_2 + 6x_3 + 7x_4 + 0x_5 + 0x_6 + 0x_7;$$

$$\left.\begin{array}{l} 2x_1 + x_2 + x_3 + x_4 + x_5 = 280, \\ x_1 + x_3 + x_4 + x_6 = 80, \\ x_1 + 2x_2 + x_3 + x_7 = 250, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,7}).$$

Таблица 4

Номер итерации	БП	1	x_1	x_2	x_3	x_4	x_5	x_6	x_7
0	x_5	280	2	1	1	1	1	0	0
	x_6	80	1	0	1	$\boxed{1}$	0	1	0
	x_7	250	1	2	1	0	0	0	1
	Z	0	−4	−3	−6	−7	0	0	0
1	x_5	200	1	1	0	1	1	−1	0
	x_4	80	1	0	1	1	0	1	0
	x_7	250	1	$\boxed{2}$	1	0	0	0	1
	Z	560	3	−3	1	0	0	7	0
2	x_5	75	0,5	0	−0,5	0	1	−1	−0,5
	x_4	80	1	0	1	1	0	1	0
	x_2	125	0,5	1	0,5	0	0	0	0,5
	Z	935	4,5	0	2,5	0	0	7	1,5

После второй итерации получили оптимальный план в базисе $\{x_2, x_4, x_5\}$. Этому базису соответствует матрица

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

В соответствии с изложенной методикой на основе табл. 4 (итерация 2) составляем матрицу

$$\begin{array}{cccccc} x_2 & x_4 & x_5 & x_5 & x_6 & x_7 \\ \end{array}$$
$$\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & -1 & -0,5 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0,5 \end{array}\right], \qquad (14)$$

в которой порядок следования единичных столбцов x_2, x_4 и x_5 определяется порядком столбцов в матрице A, а порядок следования столбцов x_5, x_6 и x_7 — порядком единичных столбцов в единичной матрице начального базиса (см. итерацию 0).

Чтобы слева от вертикальной черты образовалась единичная подматрица, переставим первую и третью строки:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0,5 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & -0,5 \end{array}\right].$$

Справа от вертикальной черты получилась обратная матрица

$$A^{-1} = \left[\begin{array}{ccc} 0 & 0 & 0,5 \\ 0 & 1 & 0 \\ 1 & -1 & -0,5 \end{array}\right].$$

Если решение ЗЛП выполняется с использованием жордановых исключений (см. начальную таблицу 5 и конечную таблицу 6), то в соответствии с изложенной методикой составляем матрицу из столбцов,

Таблица 5

БП	1	СП			
		$-x_1$	$-x_2$	$-x_3$	$-x_4$
$x_5 =$	280	2	1	1	1
$x_6 =$	80	1	0	1	1
$x_7 =$	250	1	2	1	0
$Z =$	0	−4	−3	−6	−7

Таблица 6

БП	1	СП			
		$-x_1$	$-x_7$	$-x_3$	$-x_6$
$x_5 =$	75	0,5	−0,5	−0,5	−1
$x_4 =$	80	1	0	1	1
$x_2 =$	125	0,5	0,5	0,5	0
$Z =$	935	4,5	1,5	2,5	7

в которые преобразовались столбцы матрицы A; ими стали единичные столбцы базисных переменных x_2, x_4 и x_5 (эти столбцы в табл. 6 в явном виде не записаны, и мы должны реконструировать их так, как об этом рассказывалось выше), и столбцов, в которые преобразовались единичные столбцы переменных начального базиса x_5, x_6 и x_7 (столбцы x_6 и x_7 в табл. 6 имеются, а столбец x_5 реконструируем). В результате получаем матрицу

$$\begin{array}{cccccc} x_2 & x_4 & x_5 & x_5 & x_6 & x_7 \\ \end{array}$$
$$\left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & -1 & -0,5 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0,5 \end{array}\right],$$

совпадающую с матрицей (14). Матрица A^{-1} находится после перестановки первой и третьей строк.

Пример 3. Найти матрицу, обратную для матрицы A базиса оптимального плана следующей ЗЛП:

$$\min Z = x_1 + x_2 - x_3 - 2x_5;$$

$$\left.\begin{array}{rcrcrcrcl} x_1 & - & 2x_2 & + & x_4 & & & = & -3, \\ & & x_3 & - & 2x_4 & & & = & 2, \\ & & 3x_2 & - & x_4 & + & x_5 & \leqslant & 5, \\ & & x_2 & & & + & x_5 & \geqslant & -3, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,5}).$$

Решение. Вводя дополнительные переменные x_6, x_7 и искусственную переменную x_8, составляем M-задачу: минимизировать функцию $F = (1+M)x_1 + (1-2M)x_2 + (-2+M)x_4 - 2x_5 + (-2+3M)$ при ограничениях

$$\left.\begin{array}{l} -x_1 + 2x_2 - x_4 + x_8 = 3, \\ x_3 - 2x_4 = 2, \\ 3x_2 - x_4 + x_5 + x_6 = 5, \\ -x_2 - x_5 + x_7 = 3, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,8}).$$

Решение задачи приведено в табл. 7.

В оптимальный план задачи входят переменные x_2, x_3, x_4 и x_7. Коэффициенты при них образуют матрицу A:

$$A = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 3 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

Для нахождения матрицы A^{-1} составляем матрицу, выписывая в левую ее часть преобразованные в единичные столбцы матрицы A (см. итерацию 3) x_2, x_3, x_4 и x_7, а в правую часть — преобразованные столбцы единичной матрицы начального базиса x_8, x_3, x_6 и x_7. Получаем

$$\begin{array}{cccccccc} x_2 & x_3 & x_4 & x_7 & x_8 & x_3 & x_6 & x_7 \end{array}$$

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & -6 & 1 & 4 & 0 \\ 0 & 0 & 1 & 0 & -3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 & 1 \end{array}\right]. \qquad (15)$$

Таблица 7

Номер итерации	БП	1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
0	x_8	3	−1	$\boxed{2}$	0	−1	0	0	0	1
	x_3	2	0	0	1	−2	0	0	0	0
	x_6	5	0	3	0	−1	1	1	0	0
	x_7	3	0	−1	0	0	−1	0	1	0
	F	$-2+3M$	$-1-M$	$-1+2M$	0	$2-M$	2	0	0	0
1	x_2	3/2	−1/2	1	0	−1/2	0	0	0	1/2
	x_3	2	0	0	1	−2	0	0	0	0
	x_6	1/2	3/2	0	0	1/2	$\boxed{1}$	1	0	−3/2
	x_7	9/2	−1/2	0	0	−1/2	−1	0	1	1/2
	F	−1/2	−3/2	0	0	3/2	2	0	0	$1/2-M$
2	x_2	3/2	−1/2	1	0	−1/2	0	0	0	1/2
	x_3	2	0	0	1	−2	0	0	0	0
	x_5	1/2	3/2	0	0	$\boxed{1/2}$	1	1	0	−3/2
	x_7	5	1	0	0	0	0	1	1	−1
	F	−3/2	−9/2	0	0	1/2	0	−2	0	$7/2-M$
3	x_2	2	1	1	0	0	1	1	0	−1
	x_3	4	6	0	1	0	4	4	0	−6
	x_4	1	3	0	0	1	2	2	0	−3
	x_7	5	1	0	0	0	0	1	1	−1
	F	−2	−6	0	0	0	−1	−3	0	$5-M$

В данном случае матрица (15) имеет вид (12), а потому никаких дальнейших преобразований не требуется и сразу получаем обратную матрицу

$$A^{-1} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ -6 & 1 & 4 & 0 \\ -3 & 0 & 2 & 0 \\ -1 & 0 & 1 & 1 \end{bmatrix}.$$

Предположим теперь, что решение ЗЛП выполнялось с использованием жордановых исключений (см. начальную таблицу 8 и конечную таблицу 9).

Используя табл. 9 и 8, составляем матрицу, в которой содержатся строки искомой обратной матрицы. При этом единичные столбцы левой части составляемой матрицы реконструируем по столбцу "БП" табл. 9, а столбцы правой части выписываем из табл. 9 в порядке следования пе-

Таблица 8

БП	1	СП			
		$-x_1$	$-x_2$	$-x_4$	$-x_5$
$x_8 =$	3	−1	2	−1	0
$x_3 =$	2	0	0	−2	0
$x_6 =$	5	0	3	−1	1
$x_7 =$	3	0	−1	0	−1
$F =$	$-2+3M$	$-1-M$	$-1+2M$	$2-M$	2

Таблица 9

БП	1	СП			
		$-x_1$	$-x_8$	$-x_5$	$-x_6$
$x_2 =$	2	1	−1	1	1
$x_3 =$	4	6	−6	4	4
$x_4 =$	1	3	−3	2	2
$x_7 =$	5	1	−1	0	1
$F =$	−2	−6	$5-M$	−1	−3

ременных, образовывавших начальный базис (см. табл. 8), т. е. в порядке x_8, x_3, x_6, x_7 (столбцы x_3 и x_7 реконструируем). В результате приходим к матрице (15), а затем и к матрице A^{-1}.

Пример 4. Найти матрицу, обратную для матрицы A базиса оптимального плана следующей задачи:

$$\min Z = x_1 + x_2 + x_3 + x_4 + x_5;$$

$$\left.\begin{array}{lr} x_1 + x_2 & \geqslant 20, \\ x_1 + 2x_3 + x_4 & \geqslant 30, \\ x_2 + x_4 + 2x_5 & \geqslant 20, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,5}).$$

Решение. Вводя дополнительные x_6, x_7, x_8 и искусственные x_9, x_{10} и x_{11} переменные, приходим к следующей M-задаче: минимизировать функцию

$$F = 70M + (1-2M)x_1 + (1-2M)x_2 + (1-2M)x_3 + (1-2M)x_4 +$$
$$+(1-2M)x_5 + Mx_6 + Mx_7 + Mx_8$$

при ограничениях

$$\left.\begin{array}{lr} x_1 + x_2 - x_6 + x_9 & = 20, \\ x_1 + 2x_3 + x_4 - x_7 + x_{10} & = 30, \\ x_2 + x_4 + 2x_5 - x_8 + x_{11} & = 20, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,11}).$$

Решение задачи приведено в табл. 10, из которой видно, что в оптимальный базис входят переменные x_1, x_2 и x_4. Коэффициенты при них в модели задачи образуют матрицу

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Чтобы найти матрицу A^{-1}, обратную для A, составляем, как обычно, матрицу из единичных столбцов x_1, x_2 и x_4, являющихся результатом преобразования столбцов матрицы A (см. итерацию 3), и столбцов,

Таблица 10

Номер итерации	БП	1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
0	x_9	20	$\boxed{1}$	1	0	0	0	-1	0	0	1	0	0
	x_{10}	30	1	0	2	1	0	0	-1	0	0	1	0
	x_{11}	20	0	1	0	1	2	0	0	-1	0	0	1
	F	$70M$	$-1+2M$	$-1+2M$	$-1+2M$	$-1+2M$	$-1+2M$	$-M$	$-M$	$-M$	0	0	0
1	x_1	20	1	1	0	0	0	-1	0	0	1	0	0
	x_{10}	10	0	-1	2	$\boxed{1}$	0	1	-1	0	-1	1	0
	x_{11}	20	0	1	0	1	2	0	0	-1	0	0	1
	F	$20+30M$	0	0	$-1+2M$	$-1+2M$	$-1+2M$	$-1+M$	$-M$	$-M$	$1-2M$	0	0
2	x_1	20	1	1	0	0	0	-1	0	0	1	0	0
	x_4	10	0	-1	2	1	0	1	-1	0	-1	1	0
	x_{11}	10	0	$\boxed{2}$	-2	0	2	-1	1	-1	1	-1	1
	F	$30+10M$	0	$-1+2M$	$1-2M$	0	$-1+2M$	$-M$	$-1+M$	$-M$	0	$1-2M$	0
3	x_1	15	1	0	1	0	-1	$-1/2$	$-1/2$	$1/2$	$1/2$	$1/2$	$-1/2$
	x_4	15	0	0	1	1	1	$1/2$	$-1/2$	$-1/2$	$-1/2$	$1/2$	$1/2$
	x_2	5	0	1	-1	0	1	$-1/2$	$1/2$	$-1/2$	$1/2$	$-1/2$	$1/2$
	F	35	0	0	0	0	0	$-1/2$	$-1/2$	$-1/2$	$1/2-M$	$1/2-M$	$1/2-M$

в которые преобразовались столбцы x_9, x_{10} и x_{11} единичной матрицы начального базиса (см. итерацию 3):

$$\begin{array}{cccccc} x_1 & x_2 & x_4 & x_9 & x_{10} & x_{11} \end{array}$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/2 & 1/2 & -1/2 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 0 & 1/2 & -1/2 & 1/2 \end{array}\right]. \qquad (16)$$

Чтобы получить справа от вертикальной черты в матрице (16) обратную матрицу A^{-1}, остается переставить вторую и третью строки, после чего находим

$$A^{-1} = \begin{bmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{bmatrix}.$$

Если решение ЗЛП выполнялось с использованием жордановых исключений (см. начальную таблицу 11 и конечную таблицу 12), то единичные столбцы левой части матрицы (16) реконструируют исходя из

Таблица 11

БП	1	СП							
		$-x_1$	$-x_2$	$-x_3$	$-x_4$	$-x_5$	$-x_6$	$-x_7$	$-x_8$
$x_9 =$	20	1	1	0	0	0	-1	0	0
$x_{10} =$	30	1	0	2	1	0	0	-1	0
$x_{11} =$	20	0	1	0	1	2	0	0	-1
$F =$	$70M$	$-1+2M$	$-1+2M$	$-1+2M$	$-1+2M$	$-1+2M$	$-M$	$-M$	$-M$

Таблица 12

БП	1	СП							
		$-x_9$	$-x_{11}$	$-x_3$	$-x_{10}$	$-x_5$	$-x_6$	$-x_7$	$-x_8$
$x_1 =$	15	1/2	$-1/2$	1	1/2	-1	$-1/2$	$-1/2$	1/2
$x_4 =$	15	$-1/2$	1/2	1	1/2	1	1/2	$-1/2$	$-1/2$
$x_2 =$	5	1/2	1/2	-1	$-1/2$	1	$-1/2$	1/2	$-1/2$
$F =$	35	$1/2-M$	$1/2-M$	0	$1/2-M$	0	$-1/2$	$-1/2$	$-1/2$

состава и порядка следования базисных переменных в табл. 12; столбцы же переменных x_9, x_{10} и x_{11}, составлявших начальный базис, присутствуют в табл. 12, откуда они и выписываются в правую часть матрицы (16).

Матрица A^{-1} находится после перестановки второй и третьей строк.

Как уже отмечалось, для экономико-математического анализа оптимального плана задачи линейного программирования, связанной с наиболее рациональным использованием ограниченных ресурсов, применяются двойственные оценки. Однако все их свойства сохраняются лишь до тех пор, пока запасы b_i $(i = \overline{1,m})$ ресурсов изменяются в определенных границах, так называемых *пределах устойчивости (чувствительности) оценок*. Только в этом случае двойственные оценки точно характеризуют степень влияния указанных изменений на экстремальную величину целевой функции.

Применительно к i-му ресурсу интервал устойчивости двойственных оценок по отношению к изменению уровня этого ресурса записывают так:

$$[b_i - \Delta b_i^{\text{н}};\ b_i + \Delta b_i^{\text{в}}] \quad (i = \overline{1,m}), \tag{17}$$

где $\Delta b_i^{\text{н}}$ — нижний предел уменьшения; $\Delta b_i^{\text{в}}$ — верхний предел увеличения уровня запаса i-го ресурса. Указанные пределы находят по формулам:

$$\Delta b_i^{\text{н}} = \min_{\substack{j \\ b_{ij}>0}} \{x_j^*/b_{ij}\}, \tag{18}$$

$$\Delta b_i^{\text{в}} = \max_{\substack{j \\ b_{ij}<0}} \{x_j^*/b_{ij}\}, \tag{19}$$

где x_j^* — компоненты оптимального плана задачи; b_{ij} — элементы матрицы, обратной для матрицы базиса оптимального плана.

Обратимся к примеру производственного содержания.

Пример 5. При изготовлении продукции $П_1$, $П_2$, $П_3$ и $П_4$ используются ресурсы P_1, P_2 и P_3. Запасы ресурсов, нормы их расхода на единицу продукции и цена единицы продукции приведены в табл. 13.

Таблица 13

Вид ресурса	Запас ресурса	Норма расхода на единицу продукции			
		$П_1$	$П_2$	$П_3$	$П_4$
P_1	240	2	1	1	3
P_2	60	1	0	2	1
P_3	300	1	2	1	0
Цена		4	2	3	5

Найти интервалы устойчивости двойственных оценок по отношению к изменению каждого из используемых в производстве ресурсов.

Решение. Обозначим через x_j $(j = \overline{1,4})$ объемы выпуска продукции по видам, а через Z — общую выручку от реализации готовой продукции. Тогда экономико-математическая модель задачи запишется в виде

$$\max Z = 4x_1 + 2x_2 + 3x_3 + 5x_4;$$

$$\left.\begin{array}{l} 2x_1 + x_2 + x_3 + 3x_4 \leqslant 240, \\ x_1 + 2x_3 + x_4 \leqslant 60, \\ x_1 + 2x_2 + x_3 \leqslant 300, \end{array}\right\}$$

$$x_j \geqslant 0 \quad (j = \overline{1,4}).$$

Каноническая форма этой модели представлена табл. 14, а результат решения задачи симплекс-методом — табл. 15.

Таблица 14

БП	1	СП			
		$-x_1$	$-x_2$	$-x_3$	$-x_4$
$x_5 =$	240	2	1	1	3
$x_6 =$	60	1	0	2	1
$x_7 =$	300	1	2	1	0
$Z =$	0	−4	−2	−3	−5

Таблица 15

БП	1	СП			
		$-x_4$	$-x_5$	$-x_7$	$-x_6$
$x_1 =$	60	11/5	4/5	−2/5	−1/5
$x_2 =$	120	−4/5	−1/5	3/5	−1/5
$x_3 =$	0	−3/5	−2/5	1/5	3/5
$Z =$	480	2/5	8/5	1/5	3/5

Из табл. 15 выписываем оптимальный план $\mathbf{x}^* = (x_1^*; x_2^*; x_3^*; x_4^*; x_5^*; x_6^*; x_7^*) = (60; 120; 0; 0; 0; 0; 0)$ и соответствующее максимальное значение целевой функции $Z_{\max} = 480$.

Используя соответствие между переменными исходной и двойственной задач, находим оптимальный план двойственной задачи $\mathbf{y}^* = (y_1^*; y_2^*; y_3^*; y_4^*; y_5^*; y_6^*; y_7^*) = (8/5; 3/5; 1/5; 0; 0; 0; 2/5)$ и $f_{\min} = 480$.

Переходим к определению матрицы $A^{-1} = [b_{ij}]_{3\times 3}$, обратной для матрицы базиса оптимального плана

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

Матрица (11) в нашем случае запишется так:

$$\begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_5 & x_6 & x_7 \end{array} \\ \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right] \end{array}. \qquad (20)$$

В результате симплексных преобразований (см. табл. 15) матрица (20) приобретает вид матрицы (12):

$$\begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_5 & x_7 & x_6 \end{array} \\ \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4/5 & -1/5 & -2/5 \\ 0 & 1 & 0 & -1/5 & -1/5 & 3/5 \\ 0 & 0 & 1 & -2/5 & 3/5 & 1/5 \end{array}\right] \end{array}. \qquad (21)$$

Замечаем, что в матрице (21) слева от вертикальной черты из столбцов x_1, x_2 и x_3 образовалась единичная подматрица, поэтому нет необходимости в перестановке строк, как это было в предыдущих примерах; так что искомая матрица, обратная для матрицы оптимального базиса, имеет следующий вид:

$$A^{-1} = \begin{bmatrix} 4/5 & -1/5 & -2/5 \\ -1/5 & -1/5 & 3/5 \\ -2/5 & 3/5 & 1/5 \end{bmatrix}.$$

А теперь воспользуемся формулами (18), (19) и найдем интервал устойчивости (17) двойственных оценок сначала по отношению к ограничению по ресурсу P_1. Прежде всего вычислим нижний предел уменьшения и верхний предел увеличения:

$$\Delta b_1^{\text{н}} = \min_{\substack{1 \\ b_{11}>0}} \left\{ \frac{x_1^*}{b_{11}} \right\} = \min \left\{ \frac{60}{4/5} \right\} = 75,$$

$$\Delta b_1^{\text{в}} = \max_{\substack{j \\ b_{1j}<0}} \left\{ \frac{x_j^*}{b_{1j}} \right\} = \left| \max \left\{ \frac{120}{-1/5};\ \frac{0}{-2/5} \right\} \right| = 0.$$

В соответствии с формулой (17) искомый интервал по ресурсу P_1 принимает вид

$$[b_1 - \Delta b_1^{\text{н}};\ b_1 + \Delta b_1^{\text{в}}] = [240 - 75;\ 240 + 0] = [165;\ 240].$$

Аналогичным образом находим интервалы устойчивости оценок y_i^* по отношению к изменениям уровней ресурсов P_2 и P_3: $[60; 360]$ и $[300; 450]$.

ЛИТЕРАТУРА

1. *Адельсон-Вельский Г. М., Диниц Е. А., Карзанов А. В.* Потоковые алгоритмы. — М.: Наука, 1975.

2. *Басакер Р., Саати Т.* Конечные графы и сети. — М.: Наука, 1974.

3. *Браверман Э. М.* Математические модели планирования и управления в экономических системах. — М.: Наука, 1976.

4. *Геминтерн В. И., Каган Б. М.* Методы оптимального проектирования. — М.: Энергия, 1980.

5. *Гольштейн Е. Г., Юдин Д. Б.* Задачи линейного программирования транспортного типа. — М.: Наука, 1969.

6. *Зуховицкий С. И., Авдеева Л. И.* Линейное и выпуклое программирование. — М.: Наука, 1967.

7. *Иванилов Ю. П., Лотов А. В.* Математические методы в экономике. — М.: Наука, 1979.

8. *Кузнецов А. В., Холод Н. И.* Математическое программирование. — Мн.: Выш. шк., 1984.

9. *Кузнецов А. В., Новикова Г. И., Холод Н. И.* Сборник задач по математическому программированию. — Мн.: Выш. шк., 1985.

10. *Кузнецов А. В., Холод Н. И., Костевич Л. С.* Руководство к решению задач по математическому программированию. — Мн.: Выш. шк., 2001.

11. *Кюнци Г. П., Крелле В.* Нелинейное программирование. — М.: Сов. радио, 1965.

12. *Сакович В. А.* Исследование операций. — Мн.: Выш. шк., 1985.

13. *Сакович В. А.* Оптимальные решения экономических задач. — Мн.: Выш. шк., 1982.

14. Сборник задач и упражнений по высшей математике: Математическое программирование / А. В. Кузнецов, В. А. Сакович, Н. И. Холод и др.; Под общ. ред. А. В. Кузнецова. — Мн.: Выш. шк., 1995.

15. *Свами М., Тхуласираман К.* Графы, сети и алгоритмы. — М.: Мир, 1984.

16. *Хедли Д.* Нелинейное и динамическое программирование. — М.: Мир, 1967.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ*

*Составила Е. В. Малышева

ОГЛАВЛЕНИЕ

Альберт Васильевич КУЗНЕЦОВ
Вячеслав Андреевич САКОВИЧ
Николай Игнатьевич ХОЛОД

ВЫСШАЯ МАТЕМАТИКА
МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ
Учебник
Издание четвертое, стереотипное

ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.07.953.П.007216.04.10 от 21.04.2010 г., выдан ЦГСЭН в СПб

Издательство «ЛАНЬ»
lan@lanbook.ru; www.lanbook.com
192029, Санкт-Петербург, Общественный пер., 5.

Тел./факс: (812)412-29-35, 412-05-97, 412-92-72.
Бесплатный звонок по России: 8-800-700-40-71

Подписано в печать 16.10.12.
Бумага офсетная. Гарнитура Школьная. Формат 84×108 $^1/_{32}$.
Печать офсетная. Усл. п. л. 18,48. Тираж 1000 экз.

Заказ №   .

Отпечатано в полном соответствии с качеством предоставленных диапозитивов в ОАО «Издательско-полиграфическое предприятие «Правда Севера».
163002, г. Архангельск, пр. Новгородский, д. 32.
Тел./факс (8182) 64-14-54; www.ippps.ru